DOCUMENT RESUME

ED 172 910 PS 010 649

AUTHOR Haas, Hope J.
TITLE Philosophical Thinking in the Elementary Schools: An Evaluation of the Educational Program Philosophy for Children.
INSTITUTION Rutgers, The State Univ., N.J. Inst. for Cognitive Studies.
SPONS AGENCY National Endowment for the Humanities (NFAH), Washington, D.C.
PUB DATE 76
GRANT NEH-FS-21960-75-24
NOTE 63p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Achievement Tests; *Affective Objectives; Affective Tests; Black Students; Caucasian Students; *Children; *Cognitive Ability; Cognitive Development; Educational Research; Elementary Education; Elementary School Students; Emotional Development; Empathy; *Philosophy; *Program Effectiveness; Program Evaluation; Research; Scientific Attitudes; Spanish Americans; Surveys; *Teacher Attitudes; Teacher Workshops
IDENTIFIERS New Jersey (Newark); *Philosophy for Children Program; Texas (Denton)

ABSTRACT

This essay reports the results of a preliminary evaluation of the effects of Matthew Lipman's Philosophy for Children program on fifth and sixth graders in the first year of the program's use in two settings. The program, which involves the use of a novel about children to promote discussion, attempts to engage elementary students in philosophical debate about such issues as the bases of moral decisions and the justification of one's beliefs. The study involved multiple conditions differentiated by setting (Newark, New Jersey or Denton, Texas), by teacher preparation (attendance or non-attendance at workshops), by student ethnicity (black, Hispanic or white), and by the amount of time students spent on the program materials. Also, comparisons were made with control groups not in the program. The student factors measured included curiosity, questioning, reading, listening comprehension, logical thinking, creativity, attitudes toward experimentation, and understanding of interpersonal relations. In addition, the teachers were surveyed about their attitudes toward the program. Data are presented in 16 tables and samples of test items are appended. (BH)

Philosophical Thinking in the Elementary Schools

An Evaluation of the Educational Program

Philosophy for Children

Hope J. Haas

Institute for Cognitive Studies

Rutgers University

1976

This research was conducted through the support of the National Endowment for the Humanities Grant Number ES-21960-75-24. The author extends her appreciation to Howard Gruber for his assistance in designing the evaluation and for his thoughtful reading of the manuscript; to Robert McGee for directing the Denton experiment; to David Orden for his assistance with computer programming; and to the principals and teachers from the experimental and control schools for their cooperation.

"Philosophy for Children" is an educational program designed to promote various aspects of cognitive and affective development through exploring the world of philosophy. Matthew Lipman, originator of the program, has interpreted Dewey's statement that philosophy can be taken as a general theory of education as suggesting that the method of philosophy, with its stress on dialogue, impartiality and comprehensiveness should be an integral part of the curriculum at every stage in the educational process. Inspired by Dewey and concerned that current teaching of children was less than adequate, Lipman (1969) wrote a children's novel, "Harry Stottlemeier's Discovery", to serve as a springboard for philosophical discussion. The program is based on a respect for the processes of inquiry and analysis and a belief that the grade school child is capable of engaging in independent and meaningful thinking in subjects such as ethics, aesthetics and metaphysics, which are usually reserved for high school or college. As children read the novel, philosophical issues ranging from the treatment of people as objects to the nature of the mind, are raised and discussed. An open and accepting classroom environment is emphasized for encouraging classroom discussion. Lipman and Sharp (1975) describe the classroom environment as one in which the child can develop the courage to discuss, reason, reflect and express himself and to compare and contrast his views with those of other children. Discussion in such an environment may improve critical thinking as a result of reasoning, reflection and comparison; creativity and personal development may be enhanced through self expression, and social skills may develop through the process of communicating with one's peers.

Reading and discussion are [redacted] principle activities which characterize Philosophy for Children. The final component of the experimental program consists of a wide range of classroom activities which are provided via a teachers' manual and supplementary materials. Specially devised games and exercises are suggested for facilitating cognitive and affective development. For example, exercises for improving reasoning include logical problems through which the child learns the difference in truth value between reversing sentences such as "No pencils are pigs" and "All onions are vegetables". Sentence reversals also serve as a stimulus for creativity exercises. For example, the reversal of the sentence "All onions are vegetables" would be "All vegetables are onions". One exercise in creative writing is for the child to write an essay on what the world would be like if all vegetables were onions. Understanding of oneself and others is encouraged through activities such as analyzing the reasons behind an interpersonal crisis that occurs in the story, considering alternative courses of action the character involved could have pursued, and considering what the pupil himself would have done had he been in the same situation as the character in the story.

Reading the novel, discussion and exercises are intended to complement each other in order to facilitate achievement of program objectives. Three major aims of the program outlined by Lipman and Sharp (1975) are: 1. improve reasoning ability including perceptual inferences, logical inferences and inferences from evidence; 2. develop creativity in the form of increasing spontaneity, imaginativeness and inventiveness; and 3. personal development including self confidence, emotional maturity, general self understanding and interpersonal relations.

The program was first used in the classroom in 1970 when Lipman conducted the course for nine weeks. Jaffe (1971) reports that at the end of the pilot project experimental children demonstrated significant improvement in logical reasoning while control children showed no significant improvement. Bierman (1973) compared reading achievement scores of experimental and control children who participated in the 1970 pilot project. He concluded that the philosophy program significantly improved reading scores of students two and a half years later.

The research to be reported in this paper involves a large scale implementation of the program by public school teachers who, unlike Lipman, have not had extensive training in philosophy. Two independent experiments were conducted concurrently, one in Newark, New Jersey; the other in Denton, Texas. Twice a week Newark teachers attended workshops designed to help them teach the program. Workshops provided the teacher with background information in the history of philosophy and logic. Through workshops the teacher was trained in the art of inquiry, and was taught how to guide philosophical discussions and to evoke from students their ideas and points of view. In addition, workshops emphasized the importance of inculcating in the child habits of consistency, comprehensiveness, impartiality, giving reasons for one's beliefs and learning to size-up situations (Lipman 1975). Denton teachers did not attend workshops and had no direct assistance in implementing the program.

A comprehensive testing program was designed to assess the effectiveness of Philosophy for Children. Bloom et. al. (1956) describe the fundamental cognitive and affective objectives of education. Cognitive goals include the acquisition and retention of knowledge and the development of intel-

lectual abilities and skills such as comprehension, analysis, synthesis, evaluation and application. Affective objectives consist of changes in interest, attitudes and values, adequate adjustment and the development of appreciation.

Rather than predicting improvement in a single domain, as a result of participating in the experimental program, eight major variables were considered.

Cognitive	Affective
reading	attitudes towards experimentation
listening comprehension	understanding of interpersonal relations
curiosity	
questioning	
logical reasoning	
creativity	

Method

NEWARK EXPERIMENT

Sample

Fifth and sixth grade children from two experimental and two control schools participated in the present study. Experimental and control schools were matched for geographical location and ethnic composition. Experimental and control schools from Newark's Central Ward will be referred to as Block 1. The ethnic composition of Block 1 was virtually 100% Black. Experimental and control schools from Newark's East Ward will be referred to as Block 2. Block 2 children were approximately 85% Black and 15% Hispanic. All schools were in low socio-economic status areas within a two mile radius. A total

of 369 children participated in the evaluation; 208 were in the fifth grade, 161 were in the sixth grade. The sample was not selected at random since experimental teachers had to volunteer to conduct the program and attend workshops. Most of the Block 1 experimental teachers were somewhat familiar with the philosophy program because one of their colleagues had used it in the classroom the previous year. These teachers were interested in implementing the program in their own classrooms. Experimental teachers from Block 2 had no previous knowledge of the program. They volunteered to participate after learning about the program from their principal and from Lipman. Control classes were matched to experimental classes on the basis of reading scores. None of the control teachers knew about the Philosophy for Children program until after the evaluation study was completed.

Treatment

Treatment consisted of reading, discussion and special activities. The experimental program was conducted for approximately seventeen weeks on an average of three days a week for forty minutes per day. Thus the program lasted approximately a total of thirty-four hours. The remainder of classroom activity was based on the traditional school curriculum with the same teacher. Newark teachers covered roughly 40% of the text of Harry Stottlemeier's Discovery. Each child had his own copy of the book from which he read or followed along while others read. Philosophical discussion accounted for more than half of total program activity. (Estimate based on teachers' reports)[1]. Supplementary teaching materials such as

[1] Teachers kept daily records of their experimental classroom activities. They indicated relative amounts of time spent engaged in reading, discussion and exercises. In addition teachers specified which issues were discussed, they recorded the length of each discussion and rated the quality of the discussions in terms of students' enthusiasm and the degree to which students stuck with the topic.

suggested activities and exercises were used quite frequently. As mentioned previously, experimental teachers attended workshops twice a week. Control classes were taught with the conventional classroom curriculum.

Design

Solomon's four group design (Solomon and Lessac 1968) was employed because it provides the necessary and traditional no treatment control in addition to controlling for the effects of pretesting. Brooks and Hannah (1969) in an analysis of pretest effects on the STEP Listening Comprehension Test found that STEP pretests affected posttest scores, thus causing contamination in spite of use of alternate forms of the test. They recommend Solomon's four group design as one of the methods of controlling for pretest effects. The basic unit of the design consists of two experimental and two control groups. One experimental and one control group are both pre and post tested, while one experimental and one control group are posttested only. (See Table 1 for a model of the design). This design is utilized for each of four basic units yielding a total of sixteen classes. The four basic units are as follows: Block 1 grade 5; Block 1 grade 6; Block 2 grade 5; Block 2 grade 6. Experimental groups were randomly assigned to design conditions.

Table 1.: Model of Solomon's 4 Group Design

Group No. and Description	pretest	treatment	posttest
1 Experimental	yes	yes	yes
2 Control	yes	no	yes
3 Experimental	no	yes	yes
4 Control	no	no	yes

Evaluation Program

Evaluation consisted of both individual and group testing. With the exception of the MAT (to be discussed below) all group testing was conducted in the morning and only one group test was given each day. Four of the group tests were administered by the classroom teacher. Two were administered by the author (White female) during pretesting and by an assistant (Black female) during posttesting. Both cognitive and affective variables were assessed in the testing program. Cognitive factors evaluated include curiosity, questioning, reading, listening comprehension, logical thinking and creativity. Affective variables include interpersonal relations and attitudes towards experimentation.

Group Measures

Curiosity

Two of the Maw and Maw (1964) curiosity tests were administered by the classroom teacher. The Story Satisfaction test (SS) has a verbal stimulus (a short story) and the Picture Satisfaction test (PS) has a nonverbal stimulus (a picture). Measures reflect interest in seeking information and exploring alternative aspects of a situation. Different forms were used in pre and post testing.

Questioning

A qualitative analysis was performed on questions asked by children on the Story Satisfaction tests. Questions were classified into five categories suggested by Gall (1970). Two of these categories, analytical and creative questions, were pertinent to the present investigation. Analytical questions involve the reasoning behind events, analysis of

events, logical deductions, explanations, and analysis of motivations. Creative questions involve synthesis, speculations on outcomes that go beyond the obvious or immediate, exploration of alternative possibilities, and divergence within the constraints of the situation.

Reading

Measurement of this variable was derived from the Metropolitan Achievement Test (MAT)(1962) which was administered through the city wide school testing program. Reading grade equivalent, the statistic used, reflects perception and understanding of details, recognition of the main idea, making inferences and understanding word meanings. Scores from the Spring 1974 administration were considered "pretest" measures, while Spring 1975 scores served as posttests. Alternate forms of the intermediate level test were used.

Listening Comprehension

The Cooperative Sequential Tests of Educational Progress (STEP) Listening Comprehension Test (1956) was administered by the classroom teacher. It measures various aspects of cognitive functioning including memory, understanding content, making simple and complex inferences and drawing conclusions. Alternate forms of the level four test were used.

Logical Thinking

Four of the California Test of Mental Maturity (CTMM) (1963 long form) subtests (level two) were administered by the author and a testing assistant. Three of the subtests were nonverbal measures of the recognition of similaritie opposites and analogies. The fourth was a verbal test which measured use of inductive, deductive and transitive inferences. Pre and post measurement involved use of the same test materials.

Interpersonal Relations

Ojemann's (1955) Social Causality Test (SCT) was administered by the classrrom teacher. The SCT measures: a. the child's ability to look at social situations from another person's point of view; b. understanding of the complex nature of causality in social situations; c. the ability to suspend judgment in the absence of sufficient information. The same instrument was used during pre and post testing.

Experimental Attitude

The Pupil Situational Inventory (PSI) (Cheong 1969) explores attitudes concerning; a. blind acceptance of the word of authority; b. freedom of students to develop and explore ideas; c. rigid educational values and procedures. This instrument was administered by the author and a testing assistant. The same test was used for pre and post measurement.

Individual Testing

Three boys and three girls from each pre-post group (N=48) were selected at random and tested with Piagetian verbal and nonverbal measures of logical reasoning and with measures of creative thinking. Different materials were used for pre and post testing. Order of presentation of these three basic measures was counterbalanced within each group and children were randomly assigned to the various testing orders. All individual testing was conducted by the author. Results of individual testing will be submitted in a separate report.

Piagetian Verbal

Test items consisted of verbal absurdities selected from Piaget (1928) and the Foolish Sayings Test (Maw and Maw 1964). Half were used for the pretest

and the other half (comparable items) were used for the posttest. Testing procedure was based on the clinical method employed by Piaget (1928).

Piagetian Nonverbal

Martorano (1973) demonstrated that the tests used by Inhelder and Piaget (1958) are differentially appropriate for children of various ages. The correlations and chemicals tasks were found to be most suitable for children of the upper elementary school years. Martorano's version of the correlations problem was used as the pretest while the chemicals task served as the post test. Test procedure was based on the clinical method used by Inhelder and Piaget (1958).

Creativity

Wallach and Kogan (1965) developed a battery of verbal and nonverbal measures of creative thinking. Two of their verbal tests were employed in the present study. The Similarities Test was used as the pretest; Alternate Uses served as the posttest.

DENTON EXPERIMENT

Sample

A total of 93 fifth and sixth grade children from two schools (one experimental and one control) participated in the evaluation study. There were 45 children from the fifth grade and 48 from the sixth grade. As in Newark, children were from a low socio-economic status area and schools were matched on the basis of geographical location and ethnic composition. Approximately 80% of the children were White, 12% were Black and 8% were Hispanic. Experimental teachers volunteered to participate in the program and control classes were matched on the basis of reading scores.

Treatment

The experimental program was conducted for about five and a half weeks on an average of three days a week for thirty minutes per day. Thus the

total amount of time spent teaching the program was approximately eight hours. Denton teachers covered the entire text of Harry Stottlemeier's Discovery. Supply and distribution of materials was a major problem. Apparently there were not enough copies of "Harry" for each child to have his own, which made reading somewhat difficult. In addition, frequently the supplementary teaching materials were not available when needed. As mentioned earlier, Denton teachers did not attend workshops. Control classes were taught using the traditional classroom curriculum.

Design

Solomon's four group design was planned for the Denton experiment however the experimental and control classes which were assigned at random to the post only condition were not tested. The result was a traditional pretest-posttest control group design for both fifth and sixth grades.

Evaluation Program

Testing in Denton followed the same basic plan that was employed in Newark. Robert McGee, director of the Denton experiment, provided testers for the non-teacher administered measures.

Group Measures

With the exception of reading, all variables were tested with the same materials that were used in Newark. Reading scores were based on the reading grade equivalent from the Iowa Test of Basic Skills, the achievement test battery used by the Denton public school system.

Individual Testing

Individual testing followed the same format and involved the same materials described in the Newark experiment. Twenty-four children were tested by a female graduate student from a local university.

Results

All data was subjected to computer analysis using programs derived from the Statistical Packages for the Social Sciences (SPSS). Each dependent variable was examined independently and was analyzed according to two basic strategies. First results from each grade were analyzed separately in order to obtain a detailed representation of the main effects (Newark and Denton) and interactions (Newark only). Then a combined grade analysis was performed to test for overall treatment and grade effects.

NEWARK

Reading

A multiple regression procedure was used to analyze the reading data which was the only Newark data in which there were both pre and post measures for all experimental and control classes.

1. Treatment

Separate analysis of fifth and sixth grade data revealed no significant difference between experimental and control groups in the fifth grade. Sixth grade experimental Ss scored substantially higher than their respective controls: $F\ (1,145)= 3.737$, $p<.055$. A combined analysis of the fifth and sixth grade data revealed that experimental Ss scored significantly higher than their matched controls: $F\ (1,305)= 5.796$, $p<.025$.

2. Block

Block was not a significant factor in the fifth grade however it was significant in the sixth grade. Block 2 Ss scored higher than Block 1 Ss: $F\ (1,145)= 22.063$, $p<.0005$.

3. Design

While the design effect was not significant for the fifth grade data, it was significant for the sixth grade. Ss who were in the pretested groups scored higher than Ss in nonpretested groups: $F(1,145) = 27.467$, $p < .0005$. The "design" effect does not really reflect any difference due to the design per se. The difference here reflects the initial differences between the groups, not a pretest effect, since all Ss included in the analysis were given both pre and post tests for this variable. Significance of the design effect in this case shows the importance of recognizing a more general phenomenon that occurred in this study, a problem that could not be avoided. This problem is the pervasive confound of pretest effects and initial differences between groups assigned to the basic design conditions (pre and post test; post test only). A pure design effect could only be determined under circumstances in which the groups being compared are identical except with respect to their design. Example: Ss are randomly assigned to groups A and B and therefore groups are assumed to be equal. Both groups are subjected to the same experimental manipulation. Group A is both pre and post tested; Group B is post tested only. If the results of posttesting indicate a difference between groups A and B, that difference can be attributed to the effects of pretesting. (If there is no difference on can assume that pretesting had no effect on performance) If, however, one cannot assume that A and B are equal at the outset, a difference on the post tests could be a function of either pretest effects, initial differences between the groups, or both. While experimental and control groups in the present study were matched on the basis of reading scores, the match on this variable was not perfect, but rather the closest approximation that could be made within

the existing school and classroom structures. While one can interpret a significant design effect to measure the initial difference between the groups for reading, one cannot clearly interpret the cause of design effects for the other dependent variables since both experimental design and initial differences between groups may be contributing factors, the relative effects of which cannot be partialled out.

4. Grade

Combined analysis did not reveal a significant grade effect for reading.

The remainder of the Newark data was analyzed by using analysis of variance procedures as suggested by Campbell and Stanley (1963). Campbell and Stanley highly recommend use of the four group design, and they recommend analyzing the data by doing an analysis of variance on the post test scores, disregarding pretests except as another "treatment" coordinate. They find Solomon's suggested statistic, an analysis of variance on the gain scores, unacceptable because of the assymetrical nature of the design. While it would obviously be desirable to analyze the pretest data more fully, there is no established statistical procedure which makes use of all six sets of observations simultaneously.

The basic form of the analysis was a three way analysis of variance with treatment (experimental/control), block (1/2) and design (pre-post/post only) as the main effects. The fifth and sixth grade combined four way analysis of variance included grade (5/6) as a main effect. That analysis revealed no significant grade effects and only one significant treatment effect which will be described when the results of that variable are presented.

Curiosity: Post Test Comparisons

Story and Picture Satisfaction tests were scored for the total number of questions asked (N) and the number of different ideas contained in the questions (D) thereby yielding a total of four curiosity measures (SN, SD, PN, PD).

1. Main Effects

A. Treatment

Fifth grade experimental Ss asked significantly more questions than controls; SN: $F(1,136) = 14.880$, $p < .001$; PN: $F(1,142) = 8.798$, $p < .004$. The number of different ideas was not significant in the fifth grade for either the Picture or Story Satisfaction tests. Analysis of the sixth grade data revealed that control Ss performed better than experimental Ss on both Story Satisfaction measures: SN: $F(1,117) = 9.417$, $p < .003$; SD: $F(1,117) = 13.262$, $p < .001$. Neither of the Picture Satisfaction curiosity measures was significant in the sixth grade. Combined analysis of 5th and 6th grade data resulted in a significant treatment effect for PN. Experimental Ss scored higher than controls: $F(1,256) = 9.901$, $p < .002$.

B. Block

Both 5th and 6th grade data revealed significant block effects on all curiosity measures. In each case, Block 2 Ss scored significantly higher than Block 1 Ss.

C. Design

Design effects were not significant for any of the fifth grade curiosity variables, however all sixth grade curiosity scores showed significant effects for design. In each case Ss who were not pretested scored higher than those who were pretested.

2. Interactions

A. Treatment X Block

There were no significant TxB interactions for any of the fifth grade curiosity measures. In the sixth grade, Block 1 experimental and control Ss scores on the SD variable were approximately equal, however in Block 2 control Ss scored higher than experimental Ss on the SD variable. In the case of both Picture Satisfaction measures, Block 1 experimental Ss scored higher than their controls, while the opposite pattern of results was obtained in Block 2.

B. Treatment X Design

TxD interactions were not significant for either of the 5th grade Story Satisfaction measures, however they were significant for both Picture Satisfaction measures. Experimental Ss who were pretested performed better than pretested control Ss, while nonpretested controls performed better than nonpretested experimentals. Three out of four of the 6th grade TxD interactions were significant. Pretested control Ss produced higher SD scores than pretested experimental Ss, while the performance of nonpretested experimental and control Ss was approximately equal. Picture Satisfaction scores (PN and PD) were higher for pretested control Ss than for pretested experimental Ss while nonpretested experimental Ss scored higher than nonpretested control Ss.

C. Block X Design

In the 5th grade SN, SD and PN all demonstrated the same significant pattern of results. Block 1 Ss who were not pretested scored higher than those who were, while Block 2 Ss who were pretested scored higher than those who were n

In the sixth grade only the SN BxD interaction was significant. Block 1 nonpretested Ss scored higher than Block 1 pretested Ss while scores for Block 2 pretested and nonpretested Ss were approximatley equal.

D. Treatment X Block X Design

The three way interaction was significant in the fifth grade for both Picture Satisfaction measures. On the PN variable, Block 1 experimental Ss in both design conditions performed better than their respective controls. In Block 2, pretested experimental Ss scored higher than their controls, however nonpretested controls scored higher than nonpretested experimentals. Results of PD analysis showed that Block 1 nonpretested experimental and control Ss's scores were approximately equal, while pretested experimental Ss scored higher than their controls. In Block 2, pretested experimental Ss scored higher than pretested control Ss while the opposite pattern of results was observed for Ss in the nonpretested conditions. In the sixth grade TxBxD interactions were significant for both Story Satisfaction measures. In both cases Block 1 pretested control Ss scored higher than pretested experimental Ss while the reverse effect was obtained for Ss in the nonpretested design conditions. In Block 2, control Ss in both design conditions scored higher than experimental Ss.

Interpersonal Relations: Post Test Comparisons

1. Main Effects

A. Treatment

There was no significant difference between 5th grade experimental and control Ss on the Social Causality Test. In the sixth grade, experimental Ss scored significantly higher than control Ss: $F(1,107) = 16.340$, $p < .001$.

B. Block

The block effect was not significant in the 5th grade however it was significant in the 6th grade. Block 2 Ss scored higher than Block 1 Ss: $F(1,107) = 4.130$, $p < .042$.

C. Design

In the 5th grade the design variable approached significance, nonpretested Ss scored higher than pretested Ss. Sixth grade data did not reveal a significant design effect.

2. Interactions

A. Treatment X Block

TxB interactions were not significant in either the 5th or 6th grade.

B. Treatment X Design

The TxD interaction was not significant in the 5th grade however it was significant in the 6th grade. The performance of experimental and control pretested Ss was approximately equal, while experimental Ss who were not pretested scored higher than their respective controls: $F(1,107) = 7.970$, $p < .006$.

C. Block X Design

In the 5th grade the BxD interaction was not significant, however it was significant in the 6th grade. Block 1 pretested Ss scored higher on the measure of interpersonal relations than Block 1 nonpretested Ss, while in Block 2, nonpretested Ss scored higher than pretested Ss.

D. Treatment X Block X Design

The three way interaction was not significant in either the 5th or 6th grade.

Listening Comprehension: Post Test Comparisons

1. Main Effects

A. Treatment

An analysis of the 5th grade STEP scores revealed no significant difference between experimental and control Ss. Sixth grade experimental Ss scored significantly higher than control Ss on the listening comprehension measure: $F(1,87) = 4.528$, $p < .034$.

B. Block

In the 5th grade Block 1 Ss scored significantly higher than Block 2 Ss: $F(1,128) = 5.208$, $p < .023$. For the 6th grade data a block effect could not be determined because one block was missing two sets of scores (one experimental and one control) and an analysis of such asymmetrical data would be misleading.

C. Design

In the 5th grade nonpretested Ss performed significantly better than pretested Ss on the listening comprehension test: $F(1,128) = 3.886$, $p < .048$. Sixth grade design effects could not be validly calculated since half of the posttest data was not available.

2. Interactions

A. Treatment X Block

The TxB interaction was not significant in the 5th grade and could not be determined for the sixth grade.

B. Treatment X Design

In the 5th grade pretested experimental Ss scored significantly higher than pretested control Ss on the STEP, while nonpretested Ss showed the opposite pattern of results: $F(1,128) = 17.320$, $p < .001$. The TxD interaction could not be determined in the sixth grade.

C. Block X Design

Fifth grade Ss in Block 1 who were not pretested scored higher than Ss who were pretested. In Block 2, Ss who were pretested scored higher than those who were not pretested: $F(1,128) = 51.411$, $p < .001$. It was not possible to determine this interaction in the sixth grade.

D. Treatment X Block X Design

In the 5th grade this three way interaction was not significant and it could not be determined for the 6th grade.

Experimental Attitude: Post Test Comparisons

1. Main Effects

A. Treatment

Fifth grade experimental Ss scored substantially higher than their controls on the Pupil Situational Inventory: $F(1,140) = 3.531$, $p < .059$. There was no significant difference between experimental and control subjects in the 6th grade on this variable.

B. Block

Block was not significant as a main effect in either the 5th or 6th grade.

C. Design

Design was a significant factor in the 6th grade only. Pretested Ss scored higher than nonpretested Ss: $F(1,121) = 5.199$, $p < .023$.

2. Interactions

A. Treatment X Block

In the 5th grade the performance of Block 1 experimental and control Ss was approximately equal on the PSI, while in Block 2 experimental Ss scored significantly higher than their controls: $F(1,140) = 3.876$, $p < .048$. The interaction was not significant in the 6th grade.

B. Treatment X Design

This interaction was not significant in the 5th grade however it was significant in the 6th grade. Sixth grade pretested control Ss scored higher than pretested experimental Ss, while nonpretested experimental Ss scored higher than nonpretested controls: $F(1,121) = 4.920$, $p < .02$.

C. Block X Design

A significant BxD interaction occurred in the 5th grade on the PSI, however it was not observed in the 6th grade. Analysis of 5th grade data revealed that Block 1 Ss who were not pretested scored higher than those in Block 1 who were pretested while the opposite pattern of results was observed in Block 2: $F(1,140) = 10.399$, $p < .002$.

D. Treatment X Block X Design

The three way interaction was not significant in either the 5th or 6th grade.

Logical Thinking: Post Test Comparisons

1. Main Effects

A. Treatment

There was no significant difference between experimental and control Ss in either grade on the CTMM subtests.

B. Block

Block was not a significant main effect in either the 5th or 6th grade although it approached significance in the 6th grade, Block 2 Ss tending to score higher than Block 1 Ss.

C. Design

There was no significant design effect in the 5th grade. Sixth grade Ss who were pretested scored significantly higher than Ss who were not pretested: $F(1,111) = 15.993$, $p < .001$.

2. Interactions

A. Treatment X Block

The TxB interaction was not significant in the 5th grade but it approached significance in the sixth grade. Block 1 experimental Ss scored higher on the CTMM than their controls, while the opposite effect occurred in Block 2.

B. Treatment X Design

In the 5th grade experimental Ss who were pretested scored higher on measures of logical thinking than did their respective controls, while nonpretested control Ss scored higher than experimental Ss who were not pretested: $F(1,146)= 5.313$, $p<.021$. This interaction was not significant in the sixth grade.

C. Block X Design

Both 5th and 6th grades exhibited significant BxD interactions. In the 5th grade Block 1 Ss who were not pretested scored higher than Ss who were, while in Block 2 pretested Ss scored higher than nonpretested Ss: $F(1,146)= 18.474$, $p<.001$. Analysis of 6th grade data demonstrated that Block 1 Ss who were pretested scored higher that Block 1 Ss who were not pretested, while Block 2 Ss in both design conditions scored approximately the same: $F(1,111)= 13.447$, $p<.001$.

D. Treatment X Block X Design

Neither 5th nor 6th grades exhibited significant three way interactions.

Questioning: Post Test Comparisons

Data pertaining to this variable was subjected to qualitative as well as quantitative analysis. For the qualitative analysis the author scored all of the Story Satisfaction questions and then randomly selected 15% of the data for another rater to analyze. Eighty-eight percent agreement was

obtained. None of the main effects or interactions were significant for analytical or creative questions in either the fifth or sixth grade.

DENTON

Multiple regression analyses were performed on the Denton fifth and sixth grade pre and post test data. The grades were analyzed both separately and combined.

Reading

Reading scores of experimental and control Ss did not differ significantly for 5th or 6th graders. Combined analysis of the data demonstrated no significant treatment or grade effects.

Curiosity

Fifth grade experimental Ss scored significantly higher than their controls on each of the four curiosity measures (see Table 9). Sixth grade control Ss scored significantly higher than experimental Ss on three of the four measures (see Table 11). Results of the combined analysis reveals significant treatment and grade effects. On all four measures experimental Ss scored higher than control Ss. (see Table 13). Fifth grade Ss scored significantly higher than sixth graders on the Story Satisfaction measures, while sixth grade Ss scored significantly higher on Picture Satisfaction measures (see Table 13).

Interpersonal Relations

Fifth grade control Ss scored significantly higher on the Social Causality Test than did fifth grade experimental Ss: $F(1,40) = 4.239$, $p < .05$. Sixth grade scores were not available.

Listening Comprehension

There was no significant difference between experimental and control Ss in the fifth grade on the STEP, however a significant difference was found in the sixth grade. From pre to post testing, scores of control Ss remained essentially constant, while scores of experimental Ss decreased: $F(1,42) = 7.664$, $p < .01$. Combined analysis did not result in significant treatment or grade effects.

Experimental Attitude

Independent analysis of 5th and 6th grade data showed no significant difference between experimental and control groups. Analysis of both grades combined revealed neither significant treatment nor grade effects.

Logical Thinking

Experimental and control Ss did not differ significantly on the CTMM in either the fifth or sixth grade. Combined analysis resulted in no significant treatment effect, while the grade effect approached significance, sixth graders tending to score higher than fifth graders.

Questioning

1. Analytical

There was no significant difference between experimental and control Ss in the fifth grade in the amount of analytical questions asked. In the sixth grade, experimental Ss asked approximately the same number of analytical questions on the posttest as they did on the pretest, while their controls asked fewer analytical questions on the posttest than they did on the pretest: $F(1,30) = 4.326$, $p < .05$. In the combined analysis the treatment effect approached significance. Experimental Ss improved more than control Ss in

the use of analytical questions. The grade effect was significant; fifth grade Ss asked more analytical questions than sixth grade Ss: $F(1,39) = 4.609$, $p < .05$.

2. Creative

There was no significant difference between experimental and control Ss in either the fifth or sixth grade. Combined analysis did not reveal significant treatment or grade effects.

TEACHER EVALUATIONS

At the conclusion of the experimental program teachers were asked to fill out questionnaires from which the following results were derived. Newark results are based on evaluations submitted by six out of eight experimental teachers. Denton results are based on evaluations from all four experimental teachers, including the two whose classes were not tested. There were no fundamental differences in evaluations of teachers from tested and nontested classes.

1. To what extent do you believe the program succeeded in achieving its goals? Rate on a scale from 1 - 5: unsuccessful to successful.

   Newark Mean = 3.0 Denton Mean = 2.5

2. How interested were children in the issues in general?

   Rate 1-5: not very to very

   Newark Mean = 3.1 Denton Mean = 2.8

3. How effective were classroom discussions in general?

   Rate 1-5: not very to very.

   Newark Mean = 3.7 Denton Mean = 3.0

4. How useful were exercises in general? Rate 1 - 5: not very to very.

Newark Mean = 3.7 Denton Mean = 2.5*

* Denton teachers indicated that exercises frequently were not available.

5. How well was the level of difficulty of the material suited to your pupils? Rate 1 - 5: not very to very.

Newark Mean = 3.1 Denton Mean = 3.0

6. Would you like to participate in this program again next year?

Newark 80% yes Denton 100% no**

**Denton teachers indicated that they were irritated by the problems with distribution and supply of materials.

7. Would you encourage your colleagues to participate in this program?

Newark 100% yes Denton 50% no 50% not sure

8. Do you believe that this program offers something fundamentally different from what is already being done in the classroom?

Newark 80% yes Denton 50% yes

9. How many days a week do you think this program should be taught?

Newark Mean = 3 Denton Mean = 2

10. For how many weeks do you think the program should be taught?

Newark range = 8 to 36 (entire year) Denton range = 3 to 36 (entire year)

Some Newark teachers indicated there should be flexibility in how frequently the program should be taught and how long it should be conducte

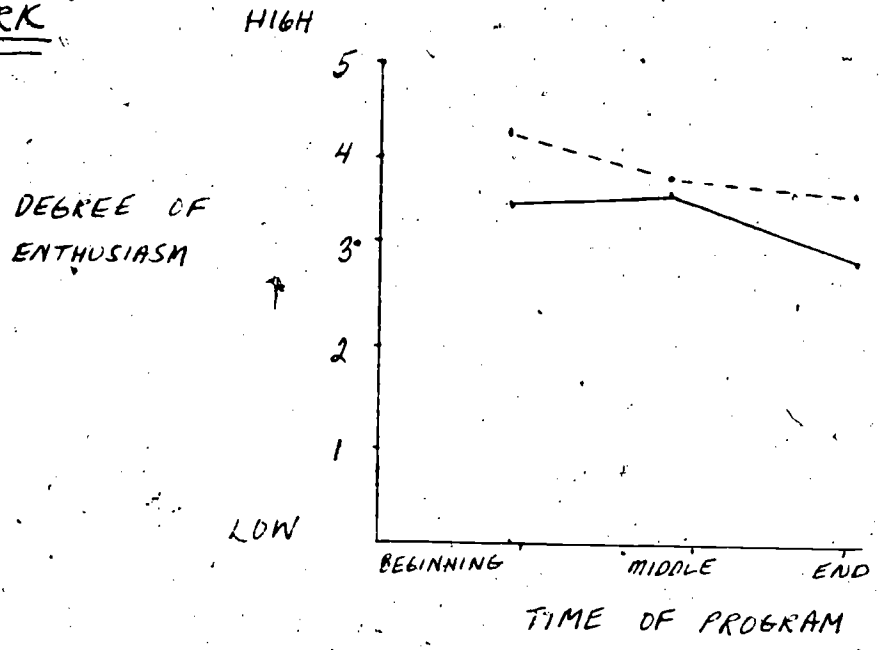
11. Newark group: How valuable do you think the workshops were in helping you teach the program? Rate 1 - 5: not very to very. Mean = 4.6

Denton group: How valuable do you think workshops held twice a week would have been in helping you teach the program?

Rate 1 - 5: not very to very. Mean = 4.5

12. Average degree of enthusiasm for the program.

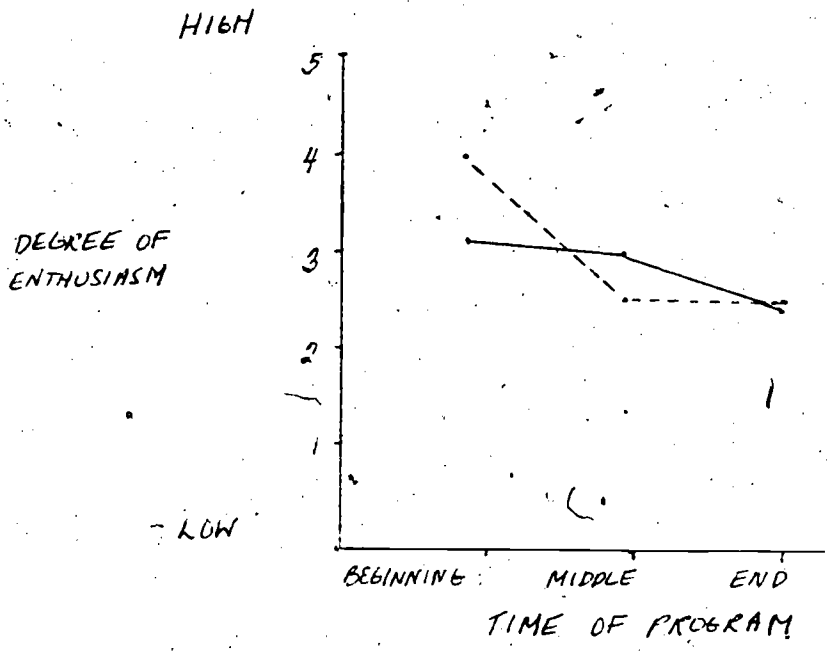
NEWARK



DENTON

TEACHER — — —

STUDENTS ———



Discussion

In Newark the experimental program resulted in significant improvement in both the cognitive and affective domains. From an educational standpoint, improvement in reading is perhaps the most important finding. As Wrightstone et. al. (1956) describe, "Reading is an essential tool for the acquisition of concepts and information in all areas of the curriculum. The ability to read with comprehension and reasonable speed contributes to the pupil's progress in all school subject matter". Overall, reading grade equivalent scores of experimental children improved eight months while control children improved five months thus resulting in a net gain of three months attributable to the four month experimental program. Improvement in reading was most dramatic for experimental children from Newark's East Ward (Block 2). Reading scores from one sixth grade class improved approximately two and a half years, while a fifth grade class improved a year and four months. The greatest gain observed in the control group was eight months and this gain was observed in one class only.

Piaget's theory of education serves as a useful model for understanding why Philosophy for Children facilitates various aspects of behavior. Piaget (1971) considers interest and action two most important pedagogical principles. Lipman, while writing the philosophical children's story, "Harry Stottlemeier's Discovery", on which the program is based, used interest as a fundamental guiding principle. In contrast to the traditional didactic textbooks elementary school children usually encounter in the classroom, Lipman (1975) describes "Harry" as a literary text that is intrinsically enjoyable and intrinsically meaningful to the child; delightful as well as instructive. Throughout the experimental program children actively read

and listen to others reading aloud from the text of this interesting novel. In addition, children engage in specially devised activities which frequently involve analyzing word meanings, recognizing the main idea contained in an episode of the story, making inferences and formulating conclusions about the material they have read. Thus Piaget's principles of interest and action characterize the reading behavior of children in the philosophy program.

While importance of reading as an educational goal is seldom underestimated, listening as an educational objective is often neglected. A survey of the most commonly used standardized achievement tests reveals that only the STEP series provides an independent measure of this fundamental skill. Yet listening behavior constitutes one of the most basic components of the educational process. Essentially the same abilities and skills which are necessary for reading comprehension are also required for listening comprehension. Both involve understanding the main idea of the material, perception of details, making inferences and drawing conclusions. Sixth grade children who participated in the experimental philosophy program demonstrated significant improvement in listening comprehension. Discussing philosohpical issues, such as the purpose of education, is one of the primary program activities. Complementary to Piaget's fundamental principles of interest and action is his belief that the important thing in an educational program is for the child to construct his own materials (Evans 1973). Philosophical discussion uniquely combines the child's interest and action, weaving them into a personal construction by encouraging him to develop and express his own ideas and to draw from his own experience. Participating in a discussion involves formulating and communicating one's ideas. The child must recognize the salient aspects of the issue being discussed, analyze their implications,

and organize them in a way that is simultaneously meaningful to himself and to others. Thus what begins as a personal construction is transformed into an interpersonal construction. While one child speaks the other children listen so that they too may participate in the discussion. Listening skills may thus be improved because of the amount of activity, level of interest and constructive aspects that characterize philosophical discussion.

Significant improvement in critical thinking in the complementary domains of reading and listening, as observed in sixth grade experimental pupils, is espeically important in light of Shapiro's reformulation of the concept of competence. She suggests that competence is not an all or none phenomenon and that the ability to shift from one mode to another may be a crucial capability. Competence, according to Shapiro (1973), involves effective functioning in different domains, the ability to respond to the requirements of different situations, flexibility in dealing with different kinds of content and in different modalities. Thus when one looks at the more general implications, the combined effect of improved comprehension in the two domains takes on greater significance than can be recognized by assessing the significance of each domain independently.

Results of the present study (Newark data) are consistent with those obtained in the pilot project in the area of reading comprehension, however they are inconsistent in the area of logical thinking. Experimental children in Lipman's pilot study demonstrated significant gains in logical thinking while experimental children in the current study showed no significant improvement. This discrepancy is probably a function of the different background and experience of the people who implemented the program in the pilot and present experiments. Lipman, a professor of philosophy, had a substantial

background in the tools and techniques of logic as well as years of experience teaching the subject on the college level. In sharp contrast, the public school teachers who conducted the program in the present experiment had little if any background in logic and virtually no experience teaching it. While workshops attempted to circumvent this problem, they were being conducted for the first time and were therefore not as effective as they potentially could be. Several teachers indicated that they did not feel adequately prepared to teach logical thinking. Since workshops were conducted concurrently with the experimental program, teachers were provided with critical background material shortly before they were to implement it in the classroom. The brief interval between training and teaching was not sufficient enough to allow teachers to feel comfortable and competent in teaching the material to their pupils. Some teachers suggested that they could have been more effective in developing logical thinking in their students if they had had more time to assimilate the material themselves.

Conflicting results on measures of curiosity were observed in both Newark and Denton. Evidence as to whether the program increases the child's level of curiosity is therefore inconclusive. Highly significant results that were obtained may be due to treatment variables or may be artifacts of the testing situation. The latter possibility seems more likely especially in light of the complex nature of the interaction effects. The instruments used are highly susceptible to extraneous influences because they are relatively unstructured. Short stories or pictures intended to provoke a child's curiosity are presented to the child who is asked what else he would like to know about them. The task is to write as many questions as one can. If a teacher were to administer the tests at the

beginning of a regular class period one could find an entirely different pattern of results than if the test were presented in the middle or at the end of a class period simply because of differential time constraints.

Evaluation of improvement in questioning behavior was based on the quantitative and qualitative analysis of questions asked on the Story Satisfaction (curiosity) test. Since the assessment of questioning was done with a measure of doubtful reliability, the results in the area of questioning must also be considered inconclusive.

Curiosity and questioning behavior was measured and analyzed in a similar manner in an educational experiment comparing self-directed study with traditional classroom instruction (lectures) for college students. Hovey, Gruber and Terrell (1963) found that students who engaged in self-directed study demonstrated improved performance on measures of curiosity and questioning when compared to students in the conventional classroom. Reliability was not a problem in the self-directed study experiment probably because the format of their test instrument (a questionnaire) was more highly structured than the format of the test materials used in the present study. The author believes that more highly standardized administration procedures could compensate for the loose format of the Maw and Maw tests to sufficiently improve reliability.

In the affective domain Newark sixth grade experimental children demonstrated increased understanding of interpersonal relations. Both the method and content of the philosophy program may have contributed to this effect. Results of the Ontario Institute for Studies in Education Moral Education Project are consistent with the finding that a program employing a discussion format can improve aspects of social development. A discussion is essentially

a social situation based on the cooperative sharing of ideas. Discussion transforms a class into a group because the children interact with each other as well as with the teacher. Through this interactive process, discussion participants are confronted with the points of view of others. Confrontation with new and conflicting ideas helps stimulate recognition of the differences between self and others. Once the difference is recognized, in order to comprehend the nature of the discrepancy more fully, the child takes the role of the other person thereby decentering from his own point of view. The transition from centering on one's own point of view to understanding the perspective of another is a major factor in social development. Sharing in each other's intellectual life through discussion also allows the child to recognize the complex reasons that underlie the ideas and actions of other people. Understanding the reasoning behind social behavior is facilitated during the process of discussion as children ask each other to explain why they feel and act as they do. The content of the experimental program helps increase understanding of social behavior because several episodes in the children's novel focus on problems that arise in interpersonal relations. One issue explored concerns treating people as objects. A brief description of an episode in the story will demonstrate how such an issue is treated. Anne, who genuinely likes and is intrigued by her new and unusual friend Suki, wants to bring Suki home to meet her parents. Suki's feelings are hurt because she thinks Anne is treating her as an interesting object to be displayed before Anne's family. Anne does not understand how Suki perceives the invitation until Suki indicates she is hurt. Suki makes a sardonic comment which shocks and confuses Anne. The comment causes Anne

to examine her own motivations. She analyzes the reasons why she wanted to bring Suki home and looks at the situation from Suki's point of view. At her moment of insight Anne says, "But Suki's a person, and you don't treat a person like a thing no matter who it is! And that's what I was doing. I was using her the way I use the cut flowers when I arrange them to make a still life. Oh, I feel awful!" ("Harry" p. 71). The combination of exploring reasons that underlie social behavior and looking at social situations from another person's point of view leads to greater understanding of the nature of interpersonal relations and to a heightened sense of social sensitivity.

Experimental children in the fifth grade did not show a significant increase in their understanding of interpersonal relations; however they demonstrated improvement which approached significance in another area of affective functioning. They developed more positive attitudes towards exploration and experimentation within an educational context. Piaget (1971) states that the experimental spirit can be most fruitfully developed beginning with the upper elementary school years when the child is coordinating his abilities to combine and dissociate factors as the formal operational structures develop. Following up Rignano's conceptualization of thinking as inner experimentation, Cronbach (1963) comments, "In a sense, 'fooling around' with indefinite aim is the essence of creativity and it may be that the key to producing more curious, more inventive adults lies in the cultivation of a playful attitude towards topics the school now treats with somber dedication to 'the right answer' ". Exploring the world of ideas is essentially the philosopher's task and as such is a fundamental aspect of this educational program. The child begins to look at familiar and accepted concepts and values in new and different ways. He questions the previously unquestioned and the

previously unquestionable in an atmosphere of tolerance and mutual respect. He has the freedom to develop and explore his ideas without worrying whether they constitute "the right answer". The nature of philosophical discussion encourages speculation and experimentation. Research evidence has demonstrated that participating in a discussion facilitates attitude change (Hovland et.al. 1953; Janis and King 1954; Mitnick and McGinnies 1958). One would therefore expect a program of this kind to lead to more positive attitudes towards intellectual experimentation.

Basically the data from the Denton experiment suggest that the program did not lead to significant improvement in either the cognitive or affective domains. Results which did reach statistical significance were difficult to interpret. As mentioned previously, equivocal findings were obtained in the area of curiosity. Other significant effects on cognitive variables (listening comprehension and analytical questions) were found in the sixth grade however they reflect relative loss rather than relative gain, and even the trends were inconsistent. The only improvement observed was in the area of interpersonal relations for fifth grade control children and the reason for this effect is abstruse.

The lack of significant improvement for experimental children in Denton could be ascribed to several factors including the duration of the program, the absence of teacher training workshops and the lack of critical instructional materials. A direct comparison of the Newark and Denton experiments is not possible because of the critical differences in the programs. However, consideration of the different pattern of results in the two cities does suggest some important implications. Duration of the program may be an important factor in determining its effectiveness.

The longer the course is conducted in the classroom the better chance there may be that it will be successful in accomplishing its goals. Teacher in both cities agreed on the desirability of workshops for implementing Philosophy for Children. Results of the teacher evaluations imply that the workshops may positively affect the teacher's attitude towards the program, the quality of classroom discussion, success or perceived success of the program, and possibly children's interest in the issues raised. Thus it appears that workshops or other forms of teacher training are obviously desirable and quite possibly effective.

Frequently when significant results are obtained from an educational experiment, critics question whether the effects obtained are truly a function of the experimental manipulation or whether they are due to the "Hawthorne Effect". A Hawthorne Effect is said to operate when special attention paid to the experimental group, causing increased morale and motivation produces the observed changes, rather than the results being due to the treatment itself. Thus there exists a potentially confounding factor within the experimental design. Substantial controversy surrounds the Hawthorne Effect. Some researchers insist on its rigorous control while others believe that the claims for its existence are grossly exaggerated. The author decided it was not necessary to include a formal control for this effect in the design of the experiment because of the improbability that special attention could lead to higher productivity on the measures employed in the current study. A post hoc analysis of the Denton experiment provides some indication that the results obtained in Newark are not a function of the Hawthorne Effect. Denton experimental children were subjected to special attention via the philosophy program for approximately five and one half weeks.

The evaluation study revealed that experimental children did not significantly improve on either cognitive or affective measures. One can therefore conclude that neither the experimental program nor special treatment operated to induce significant gains in preformance. Through these results the Denton experiment confirms that it is relatively difficult to induce change in the behavior examined in the present study. When this finding is taken into account the Newark results are not susceptible to being interpreted as a simple function of special attention; however, one cannot rule out the possibility that special attention over a longer period of time may have affected performance.

At the outset of the experiment we were not concerned with the relative effectiveness of the program in the two different grades. "Harry" was written using a fourth grade vocabulary so it was assumed that fifth and sixth grade children would benefit equally from the experimental program. Although analysis of individual dependent variables did not reveal significant differences between fifth and sixth grades for any of the variables, the overall pattern of results in Newark indicates that grade may be a factor of substantial importance. Excluding curiosity, 50% (three out of six) of the sixth grade measures resulted in significant treatment effects while only 17% (one out of six) were significant in the fifth grade. The apparently greater effectiveness of the program in the sixth grade may be a function of a higher level of intellectual development and emotional maturity which allow the older child to utilize more of what the program has to offer.

While the results of this study in terms of areas and extent of program effectiveness, grade effects, etc. are suggestive, they are certainly not

definitive. Further research needs to be conducted in order to document more conclusively the value of Philosophy for Children as an educational program. Considering that the current research was undertaken during the first year of the program's administration by previously untrained teachers, the results are most encouraging. One would expect even more significant effects to be obtained if teachers had the benefit of training and the experience of implementing the program prior to evaluation. Therefore a research program similar to the one employed here is suggested for evaluating the philosophy program's success with experienced teachers. In addition, carefully planned follow-up research should be conducted. Follow-up research can determine whether the observed gains are stable and whether they are sustained over time. Also, occasionally the effects of experience are not immediately observable. Skills, abilities or attitudes developed through the experimental program may not have noticeable effects until the gains have been consolidated through subsequent experience. Without adequate follow-up research, knowledge of delayed results would be lost. Finally, follow-up research is desirable because the effects of experience are more meaningful when viewed within a developmental context than when observed only at a single point in time.

Conclusion

An important characteristic of the experimental philosophy program is its comprehensiveness. Many educational programs, such as the Moral Education Project and the Inquiry Training Program from the Ontario Institute for Studies in Education, focus on development within a single domain. Philosophy for Children has demonstrated multifaceted development. In the realm of

cognitive functioning, experimental children in Newark significantly improved in reading and listening comprehension. In the area of affective functioning, children who participated in the program demonstrated increased understanding of interpersonal relations and more positive attitudes towards intellectual experimentation. Beyond what has already been observed, curiosity, questioning, logical thinking and creativity remain as areas of potential program effectiveness.

Abbreviations

(for reading the tables that follow)

E:	Experimental
C:	Control
T:	Treatment
B:	Block
D:	Design
SN:	Number of questions: Story Satisfaction test (curiosity)
SD:	Number of different ideas: Story Satisfaction test (curiosity)
PN:	Number of questions: Picture Satisfaction test (curiosity)
PD:	Number of different ideas: Picture Satisfaction test (curiosity)
SCT:	Social Causality Test (interpersonal relations)
STEP:	Sequential Tests of Educational Progress (listening comprehension)
PSI:	Pupil Situational Inventory (experimental attitude)
CTMM:	California Test of Mental Maturity (logical thinking)
AN:	Analytical questions
CR:	Creative questions
MAT:	Metropolitan Achievement Test (reading)
ITBS:	Iowa Test of Basic Skills (reading)

Table 2

Newark Reading Results

Separate Grade Analysis

Multiple Regression

Grade 5				Grade 6			
Variable	d.f.	F	probability	Variable	d.f.	F	probability
Treat.	1,156	2.425	n.s.	Treat.	1,145	3.737	.055
Block	1,156	2.161	n.s.	Block	1,145	22.063	.0005
Design	1,156	0.113	n.s.	Design	1,145	27.467	.0005

Reading Grade Equivalent Means

	Grade 5		Grade 6	
	Pre	Post	Pre	Post
Experimental	3.6	4.3	3.6	4.5
Control	3.7	4.1	3.5	4.1

Table 3

Newark Reading Results

Combined Grade Analysis

Multiple Regression

Variable	d.f.	F	probability
Treatment	1,305	5.796	.025
Block	1,305	14.601	.005
Design	1,305	9.227	.005
Grade	1,305	2.271	n.s.

Reading Grade Equivalent Means

	Pre	Post
Experimental	3.6	4.4
Control	3.6	4.1

Table 4

Fifth Grade: Newark

Three Way Analysis of Variance

F values and probability

Variable	d.f.	T	B	D	TxB	TxD	BxD	TxBxD
Curiosity								
SN	1,136	14.880	18.015	0.225	0.111	0.179	38.835	0.288
		.001	.001	n.s.	n.s.	n.s.	.001	n.s.
SD	1,137	2.035	10.672	2.313	0.147	0.379	9.145	0.915
		n.s.	.002	n.s.	n.s.	n.s.	.002	n.s.
PN	1,142	8.798	15.969	0.203	1.065	18.129	10.411	14.824
		.004	.001	n.s.	n.s.	.001	.002	.001
PD	1,141	0.721	8.000	0.578	0.001	10.720	2.054	5.226
		n.s.	.006	n.s.	n.s.	.002	n.s.	.02
S.C.T.	1,141	0.063	0.129	3.414	2.072	0.689	0.026	0.354
		n.s.	n.s.	.063	n.s.	n.s.	n.s.	n.s.
S.T.E.P.	1,128	0.007	5.208	3.886	2.350	17.320	51.411	2.962
		n.s.	.023	.048	n.s.	.001	.001	n.s.
P.S.I.	1,140	3.531	0.663	0.007	3.876	0.455	10.399	0.000
		.059	n.s.	n.s.	.048	n.s.	.002	n.s.
C.T.M.M.	1,146	0.084	1.888	0.001	0.737	5.313	18.474	1.790
		n.s.	n.s.	n.s.	n.s.	.021	.001	n.s.
Questions								
AN	1,93	0.009	1.506	1.970	0.882	0.145	0.106	0.563
		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
CR	1,25	1.570	0.421	1.053	0.962	0.374	0.114	2.047
		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

Table 5

Fifth Grade: Newark

Post Test Means

Variable	Experimental	Control
Curiosity		
SN	12.87**	9.26
SD	9.66	8.32
PN	9.18**	6.88
PD	6.81	6.07
Interpersonal Relations	12.17	12.38
Listening Comprehension	40.70	40.38
Experimental Attitude -	63.39*	61.47
Logical Thinking	34.57	34.26
Questions		
Analytical	3.44	3.80
Creative	1.81	1.41

**p .005

* p .059

Table 6

Sixth Grade: Newark

Three Way Analysis of Variance

F values and probability

Variable	d.f.	T	B	D	TxB	TxD	BxD	TxBxD
Curiosity	1,117	9.417	7.740	6.939	6.167	5.704	4.704	5.643
SN		.003	.006	.009	.014	.018	.03	.018
SD	1,117	13.262	8.094	4.816	7.130	7.697	6.718	4.154
		.001	.005	.028	.009	.007	.01	.041
PN	1,114	1.723	52.415	6.568	9.982	3.978	1.660	0.359
		n.s.	.001	.011	.002	.046	n.s.	n.s.
PD	1,113	0.332	39.019	6.734	12.848	5.868	1.849	1.289
		n.s.	.001	.01	.001	.016	n.s.	n.s.
S.C.T.	1,107	16.340	4.130	0.072	0.089	7.970	7.956	1.620
		.001	.042	n.s.	n.s.	.006	.006	n.s.
S.T.E.P.	1,87	4.528	-	-	-	-	-	-
		.034						
P.S.I.	1,121	0.168	0.001	5.199	0.758	4.920	1.614	0.028
		n.s.	n.s.	.023	n.s.	.02	n.s.	n.s.
C.T.M.M.	1,111	0.389	3.204	15.993	2.976	0.004	13.447	0.008
		n.s.	.073	.001	n.s.	n.s.	.001	n.s.
Questions								
AN	1,75	2.337	0.623	2.403	0.994	0.099	0.220	2.112
		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
CR	1,11	0.899	1.316	2.001	0.370	0.000	0.049	-
		n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	

Table 7
Sixth Grade: Newark
Post Test Means

Variable	Experimental	Control
Curiosity		
SN	10.57	13.90**
SD	8.19	10.73**
PN	9.50	8.28
PD	7.16	6.80
Interpersonal Relations	12.53**	10.74
Listening Comprehension	24.95*	22.63
Experimental Attitude	62.09	62.19
Logical Thinking	34.81	32.92
Questions		
Analytical	3.42	2.60
Creative	1.25	1.60

**p .005
* p .034

Table 8

Combined Grade Analysis: Newark

Four Way Analysis of Variance

Main Effects

F values and probability

Variable	d.f.	Grade	Treatment	Design	Block
Curiosity					
SN	1,253	2.237	0.474	4.122	25.715
		n.s.	n.s.	.041	.001
SD	1,254	0.419	0.293	5.047	17.555
		n.s.	n.s.	.024	.001
PN	1,256	1.128	9.901	5.178	61.345
		n.s.	.002	.022	.001
PD	1,254	0.481	1.020	4.002	29.997
		n.s.	n.s.	.044	.001
S.C.T.	1,248	1.435	1.525	2.769	0.078
		n.s.	n.s.	n.s.	n.s.
S.T.E.P.	1,215	-	0.812	-	-
			n.s.		
P.S.I.	1,261	0.173	1.835	1.393	0.515
		n.s.	n.s.	n.s.	n.s.
C.T.M.M.	1,257	0.510	0.867	7.446	4.318
		n.s.	n.s.	.007	.046
Questions					
AN	1,168	0.358	0.118	3.069	1.736
		n.s.	n.s.	.078	n.s.
CR	1,36	0.415	0.251	2.081	0.001
		n.s.	n.s.	n.s.	n.s.

Table 9

Fifth Grade: Denton

Multiple Regression

Variable	d.f.	F	probability
Curiosity			
SN	1,38	7.398	.01
SD	1,38	12.323	.005
PN	1,40	17.199	.0005
PD	1,40	11.817	.005
S.C.T.	1,40	4.239	.05
S.T.E.P.	1,38	0.110	n.s.
P.S.I.	1,36	0.720	n.s.
C.T.M.M.	1,36	0.003	n.s.
Reading	1,25	2.903	n.s.
Questions			
AN	1,8	0.523	n.s.
CR	1,2	0.041	n.s.

Table 10

Fifth Grade: Denton

Pre & Post Test Means

Variable	Experimental		Control	
	Pre	Post	Pre	Post
Curiosity				
SN	17.15	44.55**	9.81	10.67
SD	10.10	16.95***	7.19	8.29
PN	20.05	25.46***	9.43	8.38
PD	11.73	14.68***	7.48	7.33
Interpersonal Relations	15.00	14.96	17.43	18.29*
Listening Comprehension	58.00	55.70	61.68	59.47
Experimental Attitude	63.76	66.86	66.67	68.72
Logical Thinking	39.20	42.75	46.45	48.43
Questions				
Analytical	2.38	3.50	2.43	3.14
Creative	2.80	2.78	2.56	1.33
Reading	4.25	5.30	5.43	6.25

***p .005

** p .01

* p .05

Table 11

Sixth Grade: Denton

Multiple Regression

Variable	d.f.	F	probability
Curiosity			
SN	1,43	3.075	n.s.
SD	1,43	5.458	.025
PN	1,41	7.102	.025
PD	1,41	7.895	.01
S.T.E.P.	1,42	7.664	.01
P.S.I.	1,38	0.430	n.s.
C.T.M.M.	1,37	1.314	n.s.
Reading	1,34	0.000	n.s.
Questions			
AN	1,30	4.326	.05
CR	1,6	0.869	n.s.

Table 12

Sixth Grade: Denton

Pre & Post Test Means

Variable	Experimental		Control	
	Pre	Post	Pre	Post
Curiosity				
SN	9.79	12.84	17.85	23.48
SD	7.32	9.00	11.56	16.07**
PN	11.06	13.12	17.33	24.22**
PD	7.88	9.82	12.37	15.78***
Listening Comprehension	60.45	56.15	66.39	66.12***
Experimental Attitude	67.47	69.59	69.13	69.92
Logical Thinking	44.81	49.94	49.08	50.72
Questions				
Analytical	1.47	1.46*	3.54	2.96
Creative	2.00	1.17	1.46	1.92
Reading	5.87	6.71	6.67	7.54

***p .01

** p .025

* p .05

Table 13

Combined Grade Analysis: Denton

Multiple Regression

F values and probability

Variable	d.f.	Treatment	Grade	Treat.xGrade
Curiosity				
SN	1,82	12.768	15.159	10.996
		.001	.0005	.005
SD	1,82	18.683	17.525	18.011
		.0005	.0005	.0005
PN	1,82	23.328	17.207	22.845
		.0005	.0005	.0005
PD	1,82	20.000	13.642	19.499
		.0005	.0005	.0005
S.T.E.P.	1,82	3.424	3.489	4.150
		.07	.07	.05
P.S.I.	1,75	0.650	1.113	0.613
		n.s.	n.s.	n.s.
C.T.M.M.	1,75	2.231	3.480	2.409
		n.s.	.07	n.s.
Reading	1,60	0.073	1.163	0.610
		n.s.	n.s.	n.s.
Questions				
AN	1,39	3.639	4.609	3.891
		.065	.05	.06
CR	1,9	0.834	2.385	1.048
		n.s.	n.s.	n.s.

Table 14

Summary of Results

Separate Grade Analysis

Variable	Grade 5		Grade 6	
	Newark	Denton	Newark	Denton
Reading	+	+	++	--
Curiosity				
SN	++	++	--	-
SD	+	++	--	--
PN	++	++	+	--
PD	+	++	+	--
Interpersonal Relations	-	--	++	
Listening Comprehension	+	-	++	--
Experimental Attitude	++	+	-	+
Logical Thinking	+	+	+	+
Questions				
Analytical	-	+	+	++
Creative	+	+	-	-

++ Significant difference favoring experimental group

\+ Insignificant difference favoring experimental group

-- Significant difference favoring control group

\- Insignificant difference favoring control group

Table 15

Reliability of Test Instruments

Test	Type of Reliability	Coefficient
Story Satisfaction (SN)	test-retest	.75
Picture Satisfaction (PN)	test-retest	.66
Social Causality Test	unknown	.77
STEP Listening Comprehension	internal consistency	.93
Pupil Situational Inventory	test-retest	.81
CTMM Total: Short Form	unknown	.85-.89
Questioning	inter-judge agreement	.88
MAT: Reading: subtest average	unknown	.80
ITBS: Subtest estimate	split-half	.80-.89

Table 16

Group Test Administration

Test	Pre	Post
Reading	school personnel	school personnel
Curiosity	teacher	teacher
Interperaonal Relations	teacher	teacher
Listening Comprehension	teacher	teacher
Experimental Attitude	author	assistant
Logical Thinking	author	assistant

Sample Test Items

Story Satisfaction: Curiosity

John and Betty were living in a strange land. One day as they were walking home, they came suddenly upon a snake. They were very frightened and did not know which way to move because both ends of the snake looked the same. They froze in their tracks, and after what seemed like hours, although it was only a few minutes, the snake crawled away.

What else would you like to know about this story? Write as many questions as you can.

Social Causality Test: Interpersonal Relations

Henry's parents put a dollar in the cupboard. Now they are upset because Henry has taken the dollar without asking. Henry is ten years old. He has a small allowance which he may spend as he pleases.

What things do you think are important in understanding Henry's behavior?

This probably means that he can never be trusted. (True or False)

Sequential Tests of Educational Progress: Listening Comprehension

The old man hurried back to his house, and his mind was full of many things. When he suddenly saw a fat, yellow cat sitting in his best armchair, he could only stand there rubbing his eyes and wondering whose house he was in.

When the old man saw the yellow cat in his best armchair, how did he feel?

A. Pleased B. Surprised C. Sad D. Angry (Multiple Choice)

Pupil Situational Inventory: Experimental Attitude

Miss P.B. often asks her children to make guesses and try them out. Would you like a teacher like her?

YES means I like... very much

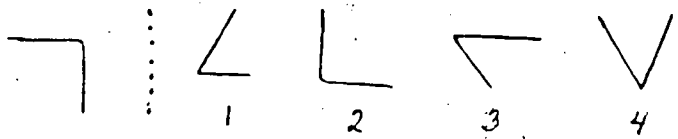
yes means I like

no means I don't like

NO means I don't like at all (Circle One)

California Test of Mental Maturity: Logical Thinking

Nonverbal: Similarities



Verbal: Inferences

Cows give milk

Goats and horses give milk

Therefore,

1. cows' milk tastes better
2. many animals give milk
3. camels give milk (Multiple Choice)

Teacher Evaluation Data

1. To what extent do you believe the program succeeded in achieving its goals?

unsuccessful	1	2	3	4	5	successful
Newark		A	B	C		
			D			
			E			
			F			
Denton		C	A			
		D	B			

2. How interested were children in the issues in general?

not very	1	2	3	4	5	very
Newark		A	B	C F		
			D			
			E			
Denton		C	A			
			B			
			D			

3. How effective were classroom discussions in general?

not very	1	2	3	4	5	very
Newark		D		A		
				B		
				C		
				E		
				F		
Denton		D	A	B		
			C			

4. How useful were exercises in general ?

not very	1	2	3	4	5	very
Newark			B	A	F	
			D	C		
			E			
Denton	C		A			
			B			
			D			

5. How well was the level of difficulty of the material suited to your pupils?

	not very 1	2	3	4	5 very
Newark		E	A	B C	
Denton		D	A C	B	

6. Would you like to participate in this program again next year?

	Yes	No
Newark	A,B,D,E	F
Denton		A,C,D

7. Would you encourage your colleagues to participate in this program?

	Yes	No	Not Sure
Newark	A,B,C,D,E,F		
Denton		C,D	A,B

8. Do you believe that this program offers something fundamentally different from what is already being done in the classroom?

	Yes	No
Newark	A,B,D,E	F
Denton	B,C	A,D

9. How many days a week do you think this program should be taught?

	1	2	3	4	5	Flexible
Newark		D	A E		F	C
Denton	B	A		C		

10. For how many weeks do you think the program should be conducted?

Newark: F: 8 weeks; D: a semester; E: entire year; A,C: Flexible

Denton: A: 3 weeks; C: $4\frac{1}{2}$ weeks; B: entire year

11. Newark: How valuable do you think workshops were in helping you teach the program?

	not very 1	2	3	4	5 very
				B C A D E	

Denton: How valuable do you think workshops held twice a week would have been in helping you teach the program?

	not very 1	2	3	4	5 very
			A		B C D

References

1. Bierman, Milton L.; A Pilot Study in the Teaching of Logic: Research Conclusions: unpublished paper Montclair Public Schools 1973

2. Bloom, Benjamin S.; et.al. Taxonomy of Educational Objectives; David McKay Company, Inc. 1956

3. Brooks, William D. and Hannah, Larry K.; Pretest Effects of the STEP Listening Test; Speech Monographs; 36; 1969 pp. 66-67

4. Campbell, Donald T. and Stanley, Julian C.; Experimental and Quasi-Experimental Designs for Research; Rand McNally College Publishing Company; 1966

5. Cheong, George S.C.; Pupil Situational Inventory: A Measure of Experimental Attitude; Journal of Experimental Education; 38(2) 1969 pp.24-30

6. Cronbach, Lee J.; Educational Psychology; Harcourt, Brace & World; 1963

7. Evans, Richard I.; Jean Piaget: The Man and His Ideas; E.P. Dutton and Company Inc. 1973

8. Gall, Meredith, D.; The Use of Questions in Teaching; Review of Educational Research; 1970, 40 pp. 707-721

9. Hovey, Donald; Gruber, Howard E.; & Terrell, Glenn; Effects of Self-Directed Study on Course Achievement, Retention and Curiosity; Journal of Educational Research; 56 (7) 1963 pp. 346-351

10. Hovland, Carl; et.al. Communication & Persuasion; Yale University Press 1953

11. Inhelder, Barbel; & Piaget, Jean; The Growth of Logical Thinking from Childhood to Adolescence; Basic Books; 1958

12. Jaffe, Jerry; Misapplication of Piaget's Developmental Model; unpublished paper; Montclair State College 1971

13. Janis, Irving & King, Bert; The Influence of Role Playing on Opinion Change; Journal of Abnormal and Social Psychology; 1954 (49) pp. 211-218

14. Lipman, Matthew; Harry Stottlemeier's Discovery; Institute for the Advancement of Philosophy for Children; Montclair State College; 1974

15. _____: Helping Children Think; unpublished paper; 1975

16. Lipman, Matthew & Sharp, Ann: Teaching Children Philosophical Thinking: Institute for the Advancement of Philosophy for Children; Montclair State College; 1975

17. Martorano, Suzanne C.; The Development of Formal Operations Thinking; unpublished Doctoral Dissertation; Rutgers University; 1973

18. Maw, Wallace; & Maw, Ethel; An Exploratory Investigation into the Measurement of Curiosity in Elementary School Children; U.S. Office of Education; Project No. 801 (SAE 8519); 1964

19. Mitnick, L. L. & McGinnies, Elliot; Influencing Ethnocentrism in Small Discussion Groups Through a Film Communication; Journal of Abnormal and Social Psychology; 56, 1958; pp. 82-90

20. Piaget, Jean; Judgment and Reasoning; Rutledge & Kegan Paul Ltd.; 1928

21. _____: Science of Education and the Psychology of the Child; Viking Press; 1971

22. Shapiro, Edna; Educational Evaluation: Rethinking the Criteria of Competence; School Review; 81 (4) 1979; pp. 523-549

23. Solomon, R.L. & Lessac, M.S.; A Control Group Design for Experimental Studies of Developmental Processes; Psychological Bulletin; 70; 1968; pp. 145-150

24. Wallach, Michael A.; & Kogan, Nathan; Modes of Thinking in Young Children; Holt, Rinehart and Winston; 1966

25. Wrightstone, J. Wayne; et.al.; Evaluation in Modern Education; American Book Company; 1956