# PHILOSOPHY FOR CHILDREN AND CALIFORNIA ACHIEVEMENT TEST: AN ANALYTIC STUDY IN A MIDWESTERN SUBURB


## ABSTRACT

This study involved 272 students in a midwestern suburban public school district. Students represented five elementary schools: three classes from each of grades two, four, and five. Experimental group students participated inthe critical thinking skills program, Philosophy for Children; control group students did not.

The following hypothesis were tested:

Hypothesis 1: Controlling for grade level and group membership, students receiving instruction in Philosophy for Children will have significantly higher gain scores on the Total Reading portion of the California Achievement Test than will students who do not receive instruction in Philosophy for Children.

Hypothesis 2: Controlling for grade level and group membership, students receiving instruction in Philosophy for Children will have significantly higher gain scores on the Total Language portion of the California Achievement Test than will students who do not receive instruction in Philosophy for Children.

Hypothesis 3: Controlling for grade level and group membership, students receiving instruction in Philosophy for Children will have significantly higher gain scores on the Total Math portion of the California Achievement Test than will students who do not receive instruction in Philosophy for Children.

Students were first assessed using analysis of variance of their 1986 California Achievement Test (CAT) scores to determine initial equivalence. At the end of the program, students' gain scores on the CAT (1987-1986 scores) in Total Reading, Total Language, and Total Math were examined using analysis of variance to determine the effects of the Philosophy for Children program on the experimental group's achievement test scores. Means for experimental group classes were higher in seven of the nine classes; analysis of variance indicated significant difference attributable to Philosophy for Children in different areas for each grade. In grade four, the Total Reading score for the group was significant at $p>.01$, and the Total Language score was significant at $p>.01$. There was no statistically significant differences for grade two.

Although the program did have some significant positive effect on students' test scores, future research is recommended to determine more specifically how Philosophy for Chilren affects achievement test scores at different grade levels in different subject areas, with more analysis regarding teacher implementation of the program.


## THE PROBLEM

Educators in the 1980s have been faced with many criticisms of public schools in a variety of reports, articles, and books. Some of the most influential of these reports are summarized by Beineke (1985) as including The Humanities in American Life: Report of the Commission on the Humanities (1980), The National Commitssion on Excellence in Education report A Nation at Risk (1983), Ernest Boyer's (1983) High School: A Report on Secondary Education in America, Adler's (1982) The Paideia Proposal: An Educational Manifesto, Goodlad's (1984) A Place Called School: Prospects for the Future. One reaction to the criticisms has been the re-emphasis of outcomes-based education; that is, to evaluate the effects of current teaching practices and the curriculum by measuring student's achievement scores on nationally standardized tests.

One of the problems that educators face in making decisions regarding curriculae adn methodology is what aspect(s) to target for change--what to add or delete in order to enhance students' achievement scores, for ultimately, schools are judged as effective or ineffective by the increase or decrease in nationally standardized test scores. According

to Squires, Huitt, Segars (1984), the literature in effective school research is particularly strong in advocating the use of standardized tests. With this focus in mind, California Achievement Test scores rather than scores on a specific test of reasoning such as The New Jersey Test of Reasoning were used for a test of program effectiveness appropriate for public schools in this particular study.

In examining the literature regarding efforts to measurably increase student achievement, one term which frequently occurs in "transfer" [Gage & Berliner (1984), Biehler & Snowman (1982), Gagne (1970), Bloom (1961), Bruner (1971), Ennis (1985)]. The premise is that skills taught in one subject area of the curriculum should assist students in other subject areas. A number of educators, including Sternberg (1986), Costa (1985), Lipman (1985), Postman (1985), Ennis (1985), and Berliner (1985), insist on the need for the inclusion of the teaching of critical thinking (reasoning) skills within the curriculum to enhance transfer. Glaser (1985, p. 24) identifies the study of critical thinking as including "skill in application," which appears to refer to one aspect of transfer.

> In the document A Nation at Risk (1983, p. 5) reference is made to the fact that . . . all children by virtue of their own efforts, competently guided, can hope to attain the mature and informed judgement needed to secure gainful employment and to manage their own lives, thereby serving not only their own interests but also the progress of society itself . . . many 17-year-olds do not possess the 'higher order' intellectual skills we should expect of them . . . nearly 40 percent cannot draw inferences from written material; only one-fifth can write a persuasive essay; and only one-third can solve a mathematics problem requiring several steps.

Another passage in A Nation at Risk (1983, p. 6) makes reference to

> . . . a problem, if schools emphasize such rudiments as reading and computation at the expense of other essential skills such as comprehension, analysis, solving problems, and drawing conclusions . . . an over-emphasis on technical and occupational skills will leave little time for studying the arts and humanities that so enrich daily life, help maintain civility, and develop a sense of community. Knowledge of the humanities . . . must be harnessed to science and technology . . . to remain relevant to the human condition.

The message appears to be that skills should not be taught in isolation, but rather integrated within a balanced curriculum. Ennis (1985, p. 44) refers to the 1980 Rockefeller Commission on the Humanities (1980) recommendation that the U.S. Office of Education include critical thinking in its definition of the basic skills in order to enhance student achievement in "the regular school curriculum."

Applebee, Langer, and Mullis (1987, p. 2) in the document Learning to be Literate in America: Reading, Writing, and Reasoning, contend that literacy is "the ability to read and write, and to reason about what one reads or writes." In commenting on that report, Howard (1987, p. 23) adds that "without reasoning in the act of reading, there is no making sense of complicated literary and informational material . . . schools, committed to an obsolete conception of literacy, have been fighting a rear-guard action that prevents the young from cultivating that function of the mind called "thinking."

According to Robert Sternberg (Critical Thinking: Its Nature, Measurement, and Improvement, 1985, p. 45), there is "an unusual amount of consensus among educators regarding the importance of including critical thinking skills in all aspects of curriculum." Critical thinking, according to Sternberg (1985, p. 46), "comprises the mental processes,

strategies, and representations people use to solve problems, make decisions, and learn new concepts." One program which purports to do this is Matthew Lipman's Philosophy for Children. In all phases of this program, students are provided with explicit models of how to bridge the gap (transfer) between critical thinking skills and everyday life (Sternberg, 1985, p. 55). Philosophy, as one area of the humanities called for in A Nation at Risk, seems a very logical, plausible approach to the inclusion of technical, foundational reasoning skills in a context which is understandable, palatable, and within the content of all levels of school curriculum. Philosophy for Children, according to Sternberg (1986, p. 19), "is appropriate for students of average, above-average, and gifted intellectual abilities," and "has transger built into the program." The Philosophy for Children program consists of a sequence of novels which are designed to introduce basic reasoning in formal and informal logic which students can use to apply to more specific problems in other subjects, as well as everyday life (Lipman, 1980, p. 15). This appears to refer to the same type of transfer of skills called for by many educators.

## DESCRIPTION OF MATERIALS

In the Philosophy for Children program, the texts which students read are called novels. Each novel is approximately the size of a slightly thicker than average magazine. The teachers' manuals are in loose leaf notebook format and are approximately four times the size of students' books. The most elementary level novel used in this study was Kio and Gus, with the accompanying teacher's manual, Wondering at the World. Kio and Gus is designed to introduce inquiry into science and environmental education, as well as the relationship between language and the world (Lipman, 1985). The target grades for Kio and Gus are two and three. The next novel in the sequence, Pixie, with the teacher's manual Looking for Meaning is targeted for grades three and four and is designed to increase the development of children's reasoning and inquiry skills such as generalization, classification, concept development, making comparisons, offering counterexamples, using analogies, contradiction and seriation (Lipman, 1985). In this study, Pixie was used by fourth grade classes, and the beginning of one fifth grade class. The third novel in the sequence is Harry Stottlemeier's Discovery with the teacher's manual Philosophical Inquiry. Harry is targeted for grades five and six. It is designed as a "non-authoritarian and anti-indoctrinational model which encourages the development of alternative modes of thought and imagination, and suggests how students are able to learn from one another" according to Lipman (1985, p. 34). In this study, one fifth grade class used only Harry, and the other fifth grade class used Harry after completing Pixie.

In summary, the aim of the program is to provide the basic reasoning techniques of critical thinking (Lipman, 1985).

## HYPOTHESES

1. Controlling for grade level and group membership, students receiving instruction in Philosophy for Children will have significantly higher gain scores on the Total Reading portion of the California Achievement Test than will students who do not receive instruction in Philosophy for Children.

2. Controlling for grade level and group membership, student receiving instruction in Philosophy for Children will have significantly higher gain scores on the Total Language portion of the California Achievement Test than will students who do not receive instruction in Philosophy for Children.

In all cases, the $p<.05$ level of significance will be used to determine whether the null hypothesis can be rejected. In addition, Omega Square will be performed on all significant F values to determine the amount of variance in the dependent variable accounted for by the independent variable.

LIMITATIONS

The following areas were identified as possible areas affecting the students' California Achievement Test scores:

1. Teachers were self-selected for participation in this project, therefore the Hawthorne Effect may be a factor in the results.
2. The program is teacher sensitive; there may be differences due to teachers' prior education in philosophy and/or other inquiry methods of teaching.
3. Teacher implement the program at different times as suits their daily schedules: all teachers may not finish the program at the same level; teachers will use different texts according to their students' reading levels; teachers will use different activities in the teacher's manual depending on their interpretation of students' responses to text discussions.

THE SAMPLE

The public school district in which this study took place is located within the metropolitan St. Louis area. The socio-economic range is from lower middle to upper middle class.

Data analysis of experimental and control groups was conducted using scores for only those students who had participated in both the 1986 and 1987 California Achievement Test (CAT) and for the experiemntal group, only those who participated in the Philosophy for Children program consistently from September, 1986 through April, 1987. Teachers were self-selected for participation in the experimental group, as recommended by the originator of the program. District administrators identified control groups by grade level in schools of similar demographics within the district. The students in the experimental group received instruction in Philosophy for Children; the control group did not.

Not all students participating in Philosophy for Children in these five elementary schools were included in the study. Third grade students district-wide did not participate in the 1987 California Achievement Test due to a new state-mandated achievement test for that grade. District administrators chose to drop the CAT for third grades for 1987 to avoide "overtesting" that grade. One fourth grade teacher and one fifth grade teacher dropped the program in the first semester because of difficulties in general with class management. Their students' scores ont he 1986 and the 1987 CAT were not used in data analysis.

DATA ANALYSIS

By inspection of Table 1, one can see that the gain scores (1987 CAT totals-1986 totals) of control and experimental groups by grade level revealed that the means of the experimental groups were higher than the means of the control groups with the exceptions: Total Reading (T-R) and Total Math (T-M) scores for grade two.

TABLE 1

Gain Scores Means (1987-1986 raw scores)
for the California Achievement Test

| Variable | N | Mean | Standard Deviation | Variance |
|---|---|---|---|---|
| | | Grade 2 Experimental Group | | |
| T-R* | 41 | 16.26829268 | 19.59977601 | 384.15121951 |
| T-L* | 41 | 56.85365854 | 22.69973676 | 515.27804878 |
| T-M* | 41 | 20.17073171 | 22.50211372 | 506.34512195 |
| | | Grade 2 Control Group | | |
| T-R | 36 | 19.63888889 | 18.28840816 | 334.46587302 |
| T-L | 36 | 47.41666667 | 23.99925594 | 575.96428571 |
| T-M | 36 | 25.72222222 | 20.36328785 | 414.66349206 |
| | | Grade 4 Experimental Group | | |
| T-R | 60 | 5.68333333 | 19.28246034 | 371.81327684 |
| T-L | 60 | 34.53333333 | 20.38732297 | 415.64293785 |
| T-M | 60 | 0.68333333 | 17.78543650 | 316.32175141 |
| | | Grade 4 Control Group | | |
| T-R | 68 | -3.77941176 | 21.32257271 | 454.65210711 |
| T-L | 68 | 25.33823529 | 22.13702935 | 490.04806848 |
| T-M | 68 | -1.39705882 | 18.04152731 | 325.49670764 |
| | | Grade 5 Experimental Group | | |
| T-R | 38 | 11.63157895 | 19.99516300 | 390.80654339 |
| T-L | 38 | 13.97368421 | 22.00305488 | 484.13442390 |
| T-M | 38 | 6.89473684 | 17.98015343 | 323.28591750 |
| | | Grade 5 Control Group | | |
| T-R | 29 | 5.93103448 | 22.14389796 | 490.35221678 |
| T-L | 29 | 12.51724138 | 21.41364833 | 458.54433498 |
| T-M | 29 | -7.51724138 | 19.56239229 | 382.68719212 |

*T-R=Total Reading Score
*T-L=Total Language Score
*T-M=Total Math Score

Although the group means were higher for the experimental groups with only two exceptions, analysis of variance revealed less differences when analyzed at the p<.05 level (see Table 2 following).

TABLE 2

One Way Analysis of Variance of Gain Scores (1987-1986 CAT) for Grade Two (note: n.s. means not significant at p<.05)

**Grade two** Dependent Variable: Total Reading

| Source | DF | Sum of Square | Mean Square | F Value |
|---|---|---|---|---|
| Model | 1 | 217.77553409 | 217.77554409 | 0.60(n.s.) |
| Error | 75 | 27072.35433604 | 360.96472448 | |
| Corrected Total | 76 | 27290.12987013 | | |

**Grade two** Dependent Variable: Total Language

| Source | DF | Sum of Square | Mean Square | F Value |
|---|---|---|---|---|
| Model | 1 | 1707.11506177 | 1707.11506177 | 3.14(n.s.) |
| Error | 75 | 40769.87195122 | 543.59829268 | |
| Corrected Total | 76 | 42476.98701290 | | |

**Grade two** Dependent Variable: Total Math

| Source | DF | Sum of Square | Mean Square | F Value |
|---|---|---|---|---|
| Model | 1 | 590.76510752 | 590.76510752 | 1.27(n.s.) |
| Error | 75 | 34767,02710027 | 463.56036134 | |
| Corrected Total | 76 | 3537.79220779 | | |

---

For grade four, group means (classes combined) of gain scores were higher for the experimental groups in all three areas: Total Reading, Total Language, Total Math; analysis of variance revealed significant differences in total Reading (significant at the p<.01 level), and in Total Language (significant at the p<.05 level) (see Table 3 following).

TABLE 3

One Way Analysis of Variance of Gain Scores (1987-1986 CAT) for Grade Four

**Grade four** Dependent Variable: Total Reading

| Source | DF | Sum of Square | Mean Square | F Value |
|---|---|---|---|---|
| Model | 1 | 2854.20049020 | 2854.20049020 | 6.86* |
| Error | 126 | 52398.67450980 | 415.86249611 | |
| Corrected Total | 127 | 55252.87500000 | | |

**Grade four** Dependent Variable: Total Language

| Source | DF | Sum of Square | Mean Square | F Value |
|---|---|---|---|---|
| Model | 1 | 2695.03576593 | 2695.02576593 | 5.92(*2) |
| Error | 126 | 57356.15392157 | 455.20757081 | |
| Corrected Total | 127 | 60051.17968750 | | |

**Grade four** Dependent Variable: Total Math

| Source | DF | Sum of Square | Mean Square | F Value |
|---|---|---|---|---|
| Model | 1 | 137.95600490 | 137.95600490 | 0.43(n.s.) |
| Error | 126 | 40471.26274510 | 321.20049798 | |
| Corrected Total | 127 | 40609.21875000 | | |

(1*) Significant at the p<.01 level.
(2*) Significant at the p<.05 level.

In summary, analysis of variance of fourth grade gain scores by group indicated significant differences attributable to Philosophy for Children for Total Reading and Total Language scores, but indicated no significant difference in Total Math scores.

For grade five, group means were higher in all three areas (see Table 1), but analysis of variance indicated significant differences only for Total Math scores (see Table 4 following).

TABLE 4
One Way Analysis of Variance of Gain Scores (1987-1986 CAT)
for Grade Five

**Grade five** Dependent Variable: Total Reading

| Source | DF | Sum of Square | Mean Square | F Value |
|---|---|---|---|---|
| Model | 1 | 534.48985562 | 534.48985562 | 1.22(n.s.) |
| Error | 65 | 28522.70417423 | 438.81083345 | |
| Corrected Total | 66 | 29057.19402985 | | |

**Grade five** Dependent Variable: Total Language

| Source | DF | Sum of Square | Mean Square | F Value |
|---|---|---|---|---|
| Model | 1 | 34.88941409 | 34.88941409 | 0.70(n.s.) |
| Error | 65 | 30752.21506352 | 473.11100098 | |
| Corrected Total | 66 | 30787.10447761 | | |

**Grade five** Dependent Variable: Total Math

| Source | DF | Sum of Square | Mean Square | F Value |
|---|---|---|---|---|
| Model | 1 | 3416.28415093 | 3416.28415093 | 9.79* |
| Error | 65 | 22676.82032668 | 348.87415887 | |
| Corrected Total | 66 | 26093.10447761 | | |

| Source | DF | ANOVA SS | F Value | PR > F |
|---|---|---|---|---|
| Group | 1 | 3416.28415093 | 9.79 | 0.0026 |

* Significant at the p<.01 level.

Secondary Analysis

Means of gain scores for individual classes indicated experimental groups were higher for T-R in 7 of the 9 classes, higher in T-L for 7 of the 9 classes, higher in T-M for 5 of the 9 classes (see Tables 5, 6, and 7). Note: In order to avoide reference to individuals' names, classes are coded. GT 21 refers to a particular grade two class, GT 22 refers to a different classroom, etc.

TABLE 5

Means for Individual Classes for Grade Two

| Variable | N | Mean | Standard Deviation | Variance |
|---|---|---|---|---|
| | | Grade two GT 21 Experimental Group | | |
| T-R | 10 | 23.00000000 | 14.83988619 | 220.22222222 |
| T-L | 10 | 65.80000000 | 14.91308151 | 222.40000000 |
| T-M | 10 | 26.10000000 | 23.77416899 | 565.21111111 |
| | | Grade two GT 21 Control Group | | |
| T-R | 18 | 18.38888889 | 19.29382373 | 372.25163399 |
| T-L | 18 | 58.05555556 | 21.96647773 | 482.52614379 |
| T-M | 18 | 35.33333333 | 13.02486310 | 169.64705882 |
| | | Grade two GT 22 Experimental Group | | |
| T-R | 17 | 12.35294118 | 19.32595786 | 373.49264706 |
| T-L | 17 | 56.82352941 | 24.50315106 | 600.40441176 |
| T-M | 17 | 21.22764706 | 21.81135700 | 475.73529412 |
| | | Grade two GT 22 Control Group | | |
| T-R | 9 | 10.11111111 | 17.67374072 | 312.36111111 |
| T-L | 9 | 30.88888889 | 15.44704215 | 238.61111111 |
| T-M | 9 | 15.22222222 | 17.49126766 | 305.94444444 |
| | | Grade two GT 23 Experimental Group | | |
| T-R | 14 | 16.21428571 | 22.69252188 | 514.95054945 |
| T-L | 14 | 50.50000000 | 24.23839929 | 587.50000000 |
| T-M | 14 | 14.78571429 | 22.84118605 | 521.71978022 |
| | | Grade two GT 23 Control Group | | |
| T-R | 9 | 31.66666667 | 9.56556323 | 91.50000000 |
| T-L | 9 | 42.66666667 | 25.85536695 | 668.50000000 |
| T-M | 9 | 17.00000000 | 27.02313823 | 730.25000000 |

TABLE 6

Means for Individual Classes for Grade Four

| | | | | |
|---|---|---|---|---|
| | | Grade four GT 41 Experimental Group | | |
| T-R | 18 | 9.27777778 | 14.50276061 | 210.33006536 |
| T-L | 18 | 36.33333333 | 18.74284177 | 351.29411765 |
| T-M | 18 | -5.38888888 | 17.40229494 | 302.83986928 |
| | | Grade four GT 41 Control Group | | |
| T-R | 17 | 13.94117647 | 10.65053161 | 113.43382353 |
| T-L | 17 | 36.41176471 | 16.86734576 | 284.50735294 |
| T-M | 17 | 7.22764706 | 12.44428761 | 154.86029412 |
| | | Grade four GT 42 Experimental Group | | |
| T-R | 11 | 8.27272727 | 18.99521471 | 360.81818182 |
| T-L | 11 | 40.27272727 | 15.00060605 | 225.01818182 |
| T-M | 11 | 1.63636364 | 14.55522399 | 211.85454545 |
| | | Grade four GT 42 Control Group | | |
| T-R | 18 | 1.44444444 | 18.24416820 | 332.84967320 |
| T-L | 18 | 29.05555556 | 15.46332368 | 239.11437908 |
| T-M | 18 | 5.00000000 | 8.22478321 | 67.64705882 |
| | | Grade four GT 44 Experimental Group | | |
| T-R | 15 | 3.73333333 | 23.60225978 | 557.06666667 |
| T-L | 15 | 34.00000000 | 24.71263413 | 610.71428571 |
| T-M | 15 | 8.80000000 | 15.26995182 | 233.17142857 |
| | | Grade four GT 44 Control Group | | |
| T-R | 18 | -14.22222222 | 18.81349814 | 353.94771242 |
| T-L | 18 | 17.22222222 | 21.33088221 | 455.00653595 |
| T-M | 18 | -6.61111111 | 18.55718947 | 344.36928105 |
| | | Grade four GT 45 Experimental Group | | |
| T-R | 16 | 1.68750000 | 20.55145980 | 422.36250000 |
| T-L | 16 | 29.06250000 | 21.36498303 | 456.46250000 |
| T-M | 16 | -0.75000000 | 20.71553362 | 429.13333333 |
| | | Grade four GT 45 Control Group | | |
| T-R | 15 | -17.60000000 | 21.23608788 | 450.97142857 |
| T-L | 15 | 18.06666667 | 29.48235954 | 869.20952381 |
| T-M | 15 | -12.46666667 | 23.94895365 | 573.55238095 |

TABLE 7

Means for Individual Classes for Grade Five

| | | | | |
|---|---|---|---|---|
| **Grade five GT 51 Experimental Group** | | | | |
| T-R | 18 | 11.61111111 | 20.17610377 | 407.07516340 |
| T-L | 18 | 11.38888889 | 21.23853016 | 451.07516340 |
| T-M | 18 | 3.88888889 | 19.16457789 | 367.28104575 |
| **Grade five GT 51 Control Group** | | | | |
| T-R | 15 | 10.13333333 | 26.20759614 | 686.83809524 |
| T-L | 15 | 14.00000000 | 23.27475641 | 541.71428571 |
| T-M | 15 | -12.26666667 | 21.27865015 | 452.78095238 |
| **Grade five GT 53 Experimental Group** | | | | |
| T-R | 20 | 11.65000000 | 20.35545963 | 414.34473684 |
| T-L | 20 | 16.30000000 | 22.96014854 | 527.16842105 |
| T-M | 20 | 9.60000000 | 16.87227561 | 284.67368421 |
| **Grade five GT 53 Control Group** | | | | |
| T-R | 14 | 1.42857143 | 16.56090989 | 274.26373626 |
| T-L | 14 | 10.92857143 | 19.97484682 | 398.99450549 |
| T-M | 14 | -2.42857143 | 16.81443650 | 282.72527473 |

In summary, the means for the individual classes for grade two were higher for the experimental groups when compared class by class with the following exceptions: class GT 21 Total Reading score; class GT 23 Total Reading and Total Math scores (16 out of 18 means were higher for the experimental groups). In grade four the means were higher for the experimental group except for GT 41 Total Reading and Total Math scores, GT 42 Total Math score (21 out of 24 means were higher for the experimental groups). In grade five, the means were higher for the experimental group except for GT 51 Total Language score (11 out of 12 means were higher for the experimental groups).

TEST HYPOTHESES

1. Controlling for grade level and group membership, students receiving instruction in Philosophy for Children will have significantly higher gain scores on the Total Reading portion of the California Achievement Test than will students who do not receive instruction in Philosohpy for Children.

Data analysis indicated that the null hypothesis was rejected for grade four ($p<.01$), but not for grades two and five. In grade two, Total Reading means for two of the three experimental classes were higher, but not enough to be significant at the >05 level of analysis.

In grade five, the group mean (the classes combined) was higher for the experimental group, but it was not significant at the $P<.05$ level.

2. Controlling for grade level and group membership, students receiving instruction in Philosophy for Children will have significantly higher gain scores on the Total Language portion of the California Achievement Test than will students who do not receive instruction in Philosophy for Children.

The null hypothesis was rejected for grade four ($p<.05$), but was not rejected for grades two and five. In grade two, two of the groups had a higher mean, and in grade five, one of the two experimental groups had a higher Total Language mean, but not enough to be significant at the .05 level.

3. Controlling for grade level and group membership, students receiving instruction in Philosophy for Children will have significantly higher gain scores on the Total Math portion of the California Achievement Test than will students who do not receive instruction in Philosophy for Children.

The null hypothesis was rejected for grade five (significant at $p<.01$), but was not rejected for grades four or two.

## SUMMARY AND IMPLICATIONS

As noted earlier, schools are being criticized by many people representing many points of view about education. One of the criticisms has been the lack of teaching skills which enable students to become better thinkers--better problem solvers--both in and out of the classroom. Those who advocate the teaching of thinking skills do so on the basis that such skills enhance students' performance throughout the curriculum. There are a variety of programs available for teaching thinking skills, some of which are summarized earlier in this article. The study discussed in this article was conducted to identify the effects of a language-based thinking skills program, Philosophy for Children, on students' standardized achievement test performance. Although the Institute for the Advancement of Philosophy for Children (IAPC, 1986) has reported various studies indicating the effectiveness of Philosophy for Children as reflected in increased achievement test scores, none of the prior studies had the same range of grades as this study.

Materials in the Philosophy for Children program are chosen by teachers according to the reading abilities of their students. Second grade classes in this study used the text Kio and Gus, fourth grade classes used Pixie; one fifth grade first completed Pixie then began Harry Stottlemeier's Discovery, and the other fifth grade used solely Harry Stottlemeier's Discovery. Each children's reader is approximately 100 pages; the teacher's manual is approximately 400 pages. Reasoning skills are developed sequentially, each level building upon the other. Reading and mathematics involve reasoning; reasoning is made up of many specific thinking skills--skills which are developed in Philosophy for Children. That both fifth grade classes had significant increases in their CAT math scores even though they used different texts at times should be significant for teachers and administrators in implementing the program, particularly in classrooms with wide ranges in students' reading abilities.

The students in this study were identified as typical of mainstream classrooms (there were some students who left the classroom at other times during the day for remedial or gifted programs). Other than for normal student absences from school, all students within the designated experimental classrooms participated in the program throughout the school year. Reading abilities of students were identified by teachers only ingeneralizations, ranging from approximately one year above grade level to one or one and one/half years below grade level. Students were not identified individually as having any handicaps such as learning or behavioral disorders, nor were students identified by race or ethnic background by the district so no data analysis was possible for those areas.

Lipman has described the program as being highly teacher-sensitive. Some of the comments made by teachers in post-program survey include:

> "The students (grade two) enjoyed the program. Initially some were hesitant to freely express their ideas. I observed much growth as the year progressed. That (increased ability to question) helped in other areas."

"I enjoyed working with the program (however) I think waiting until after Christmas to start with second graders would be better."

"I think the book is too difficult for most second graders and many of the ideas in the teacher's manual were far too complex." (Teachers are not supposed to cover the entire manual, but select only what is relevant for their class. It may be significant to note that this classroom had the lowest gain score means of the grade twos.)

"My kids love it." (grade four)

"Many skills overflowed into other areas. Their vocabulary has improved. At times Pixie's personality seemed a little immature to class (fourth grade) but they really enjoyed the program."

"What a fabulous program." (grade five)

Second grade teachers in this study made comments about their students' difficulties with reading the program in the first semester, yet means for Total Reading were higher for all experimental classes except GT 23, and for all Total Language experimental classes (see Table 3).

One fifth grade teacher assigned some of the reading as homework, contrary to the author's recommendations, reserving classroom time for discussion adn followup exercises. It may be significant to note that the teacher's students had only one point higher in Total Reading mean than did the control group, andthe Total Language mean was 3 points less thant he control group. The other fifth grade teacher's students had Total Reading mean scores nine points higher than the control and for Total Language, six points higher. However, both fifth grade experimental classes had statistically significant results ($p<.05$) in analysis of variance of gain scores in Math ont he California Achievement Test.

As indicated earlier, this program is highly teacher sensitive (IAPC, 1985). Lipman also indicates that the program is more effective when school administrators attend the training sessions along with teachers. Administrators in this case did not participate in the training.

## LIMITATIONS

Potential limitations were identified prior to the study as possibly affecting students' performance, but which could not be eliminated through controls. One of the possible limitations was that teachers would implement the program at different times during the daya nd on different days of the week and as a result, they would not all be in the same place in the text at the time of the 1987 testing. This was very likely a factor in the mixed results, and as indicated by teachers' comments ont he May, 1987 post-program survey. It was not possible to control for the ways teachers implemented the program, partially becuase the program is designed so students in each particular calss discover reasoning areas and skills in each text as their interests, not the teacher's interest, dictate and that at the time of this study, participation was voluntary, rather than a regulated portion of the curriculum. The teacher's manual contains approximately three to four times the material than can be covered by one class in a typical school year, a notable difference form a typical teacher's guide to a basal reading series. It is common to hear of educational materials being touted as "teacher proof," a term which generally refers ot very specific activites that teachers do as designed with very little or not interpretation required. This is not the case with Philosophy for Children. With a program so highly teacher-sensitive, it is impossible to control for all possible variations inherent inteaching, in dealing with different personalities and the constantly fluctuating

course of events common in school classrooms.

Another potential limitation not evidenced by the results in this study was that there did not appear to be any differences attributed to any teacher's prior training in philosophy and/or other inquiry methods of teaching. This should be a positive factor ofr other districts which may be considering implementing the program.

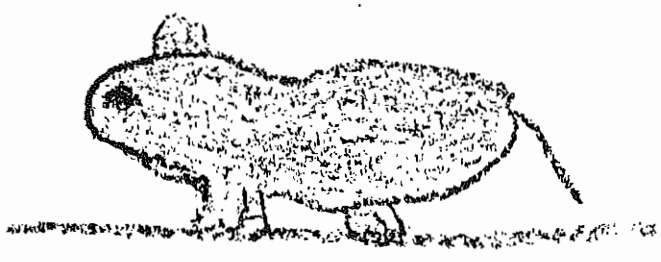
Although the variance statistically in this study was small, it appears taht this program has had some very positive effects onthe students and teachers in this study. All teachers who began the program planned to use the program again the next year, including the two who stopped it early this year due to problems with class management. AAlthough this study indicated mixed results when statistically analyzed at the $p<.05$ level, several administrators (district building prinipals) indicated taht the results were positive enough to be satisfied with the program so far, to encourage other teachers to join the programa nd ot expand into middle school-junior high. This support from administration should be very encouraging to other fledgling programs.

Another very significant factor is that the Philosophy for Children program has been identified byt he U.S. Department of Education's (1986) Joint Dissemination Review Panel as a meritorious educational program,a dn has been granted national validation bythe Department.

## RECOMMENDATIONS FOR FUTURE RESEARCH

The California Achievement Test (1986 Class Management Guide) purports to measure problem solving as well as reading comprehension. Further research with this instrument is needed to identify which subskill areas within the Total Reading, Total Language, and Total Math scores are affected most by Philosophy for Children. Second grade teachers indicated that they felt that their student's abilities at the beginning of secondgrade were not at the level required for optimum performance int he program and that perhaps it would have been more effective to delay starting the program until the seond semester. Reading levels of individual students in this study were not available. It may have been the case that certain students had wider ranges inreading ability which would have affectedthis study since it is a language-based program. IN particular, it seems curious thatthe positive scores evidenced by fourth graders in this study in reading and language scores were not evidenced for fifth graders.

Joyce Banks



Marie Acuna
Age 10