DOCUMENT RESUME

ED 202 905 TM 810 349

AUTHOR Moskowitz, Joel M.; And Others
TITLE A Process and Outcome Evaluation of a Magic Circle Primary Prevention Program.
INSTITUTION Pacific Inst. for Research and Evaluation, Napa, Calif.
SPONS AGENCY National Inst. on Drug Abuse (DHEW/PHS), Rockville, Md.
PUB DATE 29 Aug 80
NOTE 53p.; For related document, see TM 810 350.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Elementary School Teachers; Formative Evaluation; Grade 3; *Group Discussion; *Inservice Education; Primary Education; *Program Evaluation; Self Esteem; *Student Participation
IDENTIFIERS Human Development Program; *Magic Circle

ABSTRACT

In an evaluation study employing a pretest-posttest control group design, 13 elementary schools were paired based on characteristics of their student bodies, faculties, principals, and special programs. One school from each pair was randomly assigned to experimental and control conditions. Third grade teachers in the experimental schools were provided with in-service training in the Magic Circle primary prevention strategy. Process evaluation data included: (1) documentation of in-service training by outside observers; (2) teacher feedback on individual training sessions; (3) questionnaire surveys of teachers at the end of training and the school year; (4) interviews with a sample of teachers and principals; (5) observations of classroom implementation, and (6) weekly reports by teachers about classroom implementation. Process evaluation showed that the teachers mastered most of the skills taught, and they found most of them to be useful in the classroom. Student- and class-level analyses of variance/covariance on outcome data showed that boys in the experimental group had significantly higher social self-esteem at the end of the school year. No effects of Magic Circle were obtained for girls, and the level of exposure to Magic Circle was not related to experimental students' post-test outcomes. (Author/GK)

A Process and Outcome Evaluation of a Magic Circle Primary Prevention Program

Submitted to
National Institute on Drug Abuse
Prevention Branch

August 29, 1980

The Napa Project
Pacific Institute for Research and Evaluation
905 Jefferson Street
Napa, California 94558

Joel M. Moskowitz, Ph.D.
Eric Schaps, Ph.D.
Janet Malvin, Ph.D. Cand.

ABSTRACT

In an evaluation study employing a pretest-posttest control group design, 13 elementary schools were paired based on characteristics of their student bodies, faculties, principals, and special programs. One school from each pair was randomly assigned to experimental and control conditions. Fourteen third grade teachers in the six experimental schools were provided with in-service training in the Magic Circle primary prevention strategy. The teachers conducted Magic Circle in their classrooms with 217 students over the course of a school year. Fourteen third grade teachers and 250 students in the seven control schools served as a comparison group and received no interventions.

The Magic Circle in-service training consisted of ten, weekly, two-hour, after-school sessions, supplemented by a minimum of four classroom visits by the trainer to observe and guide teacher implementation of "Circles."

Process evaluation data included a) documentation of in-service training by outside observers, b) teacher feedback on the individual training sessions, c) questionnaire surveys of teachers at the end of training and the end of the school year, d) interviews with a sample of teachers and principals, e) observations of classroom implementation, and f) weekly reports by teachers about classroom implementation. Process evaluation showed that the training was very highly regarded by the teachers, that all the teachers mastered most of the skills taught in the training, and that they found most of the skills to be quite useful in the classroom. The number of Circles that teachers conducted in their classrooms varied considerably, ranging from a total of 8 per student to 36 per student, and averaging 22 per student.

Experimental and control teachers and students were pre- and post-tested, and data regarding achievement and attendance were gathered from school district records. Teachers also rated their students' classroom behavior.

Post-test comparisons of experimental and control group teachers showed that experimental teachers were significantly more satisfied with teaching than controls, and were using teaching skills related to the training significantly more often.

Student- and class-level analyses of variance and covariance on outcome data showed that compared with controls, boys in the experimental group had significantly higher social self-esteem at the end of the school year. Also, experimental boys were rated by their teachers as presenting more minor but fewer major discipline problems.

No effects of Magic Circle were obtained for girls. Furthermore, the level of exposure to Magic Circle was not related to experimental students' post-test outcomes.

Magic Circle is the popular name for the Human Development Program developed by Palomares, Bessel, and Ball (Ball, Note 1). The program was devised as a primary prevention strategy for positively influencing students' social and emotional development. According to its creators, expected effects of the program range from improved communication and listening skills to improved self-concept and peer relationships. Magic Circle is used widely; over 9,000 individuals have been trained to conduct "Circles" since 1977 (Ward, Note 2).

Magic Circle is a curriculum for conducting small group discussions in the classroom. In a Circle session, which typically lasts 15-30 minutes, the teacher leads a group of 5-12 students through a discussion of a particular topic. The teacher employs methods which encourage the students' expression of thoughts and feelings without running many of the risks of other group interaction formats. Participation is always voluntary. Confrontation and debate are never allowed. Although students may be asked about their feelings, extended probing and analysis of individual comments are strictly forbidden. The key elements of the Circle are that everyone gets an opportunity to contribute--even if they choose to remain silent--and that each contributor is listened to carefully by the other Circle members.

A number of unpublished outcome evaluations of Magic Circle have been conducted since 1970. Most of the available research has been summarized by the Human Development Training Institute (Note 3), the

developers of the program. Twenty-one of the studies summarized in the HDTI review employed experimental or non-equivalent control groups. In these studies, according to the HDTI review, preschool through grade 7 students participated in 5-150 Circle sessions. On the average, students attended more than three sessions per week ($\underline{M}$ = 3.4, $\underline{SD}$ = 1.5) over a 19 week period ($\underline{M}$ = 18.7, $\underline{SD}$ = 9.3), representing participation in a total of 60 Circle sessions ($\underline{M}$ = 60.4, $\underline{SD}$ = 51.4).

The student outcomes examined in the 21 studies included measures of five variables: self-concept, peer relationships, locus of control, reading performance, and student behavior. Twelve of these studies investigated the effect of Magic Circle upon a single variable. Only three studies assessed effects on three or more variables. Of the 12 studies that examined the effect of Magic Circle upon student self-concept, nine found no evidence for an effect, and three found positive effects for some experimental groups but not others. Four studies measured effects on peer relationships. Three of these found positive effects and the fourth found a mixed negative effect. Four studies measured locus of control, or students' feelings of responsibility for their own behavior. One of these found evidence of a positive effect. Seven studies measured effects on reading performance. Four showed positive effects, one showed mixed effects; and two showed negative or no effects. Of the seven studies in which teachers rated students' behavior, four showed positive effects, and three found no effect.

These patterns of results provide some evidence that Magic Circle can improve students' peer relationships, reading performance, class-

room behavior, and, perhaps, self-concept and locus of control. These results should be interpreted cautiously, however, since the research summaries were written by the program developers, and, more importantly, since the research suffers from several methodological shortcomings. First, previous studies have generally failed to examine systematically Magic Circle's impact upon the range of variables of interest; most studies measured only one or two variables. Better understanding of Magic Circle's effects would follow from studies which examine the various relationships among the important outcome variables.

Second, in many instances, the measures selected to operationalize outcome variables have not been specific to Magic Circle objectives. For example, general self-concept measures have been widely utilized, instead of measures more sensitive to changes in social self-concept.

Third, most outcome studies have not been accompanied by adequate process evaluation. Process evaluation provides data on the treatment implementation, e.g., the quality as well as the quantity of Circles conducted in each classroom. Such information makes possible investigations of process-outcome relationships and enables better identification of important treatment aspects.

The present study attempts to address the shortcomings described above by coupling comprehensive outcome and process evaluations of a Magic Circle teacher in-service training program. The Magic Circle training was provided in the first year of a three-year program of interventions to be delivered to a cohort of students as part of an ongoing effort to implement and evaluate promising primary prevention strategies.

The present evaluation assessed the impact of Magic Circle on the following student outcome variables: social self-esteem, attitudes toward peers, attitudes toward teacher, academic self-esteem, locus of control, attitudes toward school, perception of peers' attitudes toward school, academic achievement, school attendance, and behavior problems. Among these variables, we hypothesized that Magic Circle would be most likely to impact social self-esteem and attitudes toward peers, because the strategy most directly bears on these variables and because they have been impacted in previous research. Positive effects upon the other variables were also anticipated, although Magic Circle does not address these as directly.

Teacher outcomes examined in the evaluation included the importance of and effectiveness at implementing Magic Circle objectives, the use of Magic Circle skills in the classroom, personal satisfaction with teaching, and faculty cohesiveness. We expected that Magic Circle in-service training would positively affect these teacher outcomes with the exception of faculty cohesiveness (because the training was only provided to the grade 3 teachers at each experimental school).

METHOD

Assignment of Schools to Condition

Thirteen elementary schools (grades K-6) from a predominantly white, middle-class, suburban public school system in Northern California were paired based on characteristics of their students, faculties, principals and special programs. One school from each pair was randomly assigned to the experimental condition and the other to the control condition.[1]

Table 1 shows, for the experimental and control conditions, the school means and standard deviations of the measures employed in matching the schools. For each school, these measures included: a) total student enrollment in the spring prior to the study; b) the number of students in the cohort group during the prior spring; c) the number of "pure" grade 3 (the cohort grade) classes and the number of "mixed" classes containing grade 3 students; d) average total reading scores for the cohort group on the Stanford Achievement Test administered during the prior spring; e) combined reading comprehension and vocabulary scores for the cohort group on the Stanford Diagnostic Reading Test, administered at the beginning of the study; f) the percentage of students from families receiving Aid for Families with Dependent Children (AFDC); g) the percentage of students classified

[1]Since an odd number of schools existed, one triplet was formed from which one school was randomly assigned to the experimental condition and the other two to the control condition.

TABLE 1

COMPARISON OF EXPERIMENTAL AND CONTROL SCHOOLS AFTER MATCHING AND RANDOM ASSIGNMENT TO CONDITION

	Experimental Schools (N=6)		Control Schools (N=7)	
	School Mean	Standard Deviation	School Mean	Standard Deviation
Total School Enrollment	340	130	294	49
Cohort Grade Enrollment	51.0	24.5	40.3	11.7
Number of Cohort Classes per School				
Grade 3 only classes	2.00	0.63	1.29	0.95
"Mixed" grade classes	0.33	0.52	0.57	0.98
SAT Reading Achievement Scores	115.2	7.8	117.3	12.2
SDRT Reading Comprehension and Vocabulary Scores	66.0	4.6	65.°	6.0
Percent of Enrollment Receiving AFDC	13.0%	8.0%	11.6%	10.4%
Percent of Spanish-Speaking	2.3%	3.2%	1.5%	3.1%
Average Number of Unexcused Absences per Pupil	2.0	0.3	1.8	0.9
Socioeconomic Status Rating	46.7	34.2	47.1	30.1
Eligible Faculty Support Rating	0.86	0.22	0.93	0.19
Eligible Faculty Competence Rating	0.45	0.39	0.60	0.45
School Principal Support Rating	0.75	0.42	0.71	0.39
School Principal Competence Rating	0.66	0.41	0.29	0.39

as Spanish-speaking with either limited or no English language skills; h) the average number of unexcused absences per student during the prior school year; i) an estimate of the student body's socioeconomic status (made by a school district official).

In addition, two district officials rated each school regarding the degree to which: a) eligible teachers would support and participate in the in-service training; b) eligible teachers were already competent in classroom management and interpersonal skills; c) the principal would support the in-service training; and d) the principal had influence over his or her teaching staff.

Subjects

Initially, all grade 3 students and their teachers in the six experimental schools constituted the experimental group. One class of students and its teacher were dropped from the study because the teacher left the Magic Circle in-service training after the third session and never conducted Magic Circles in her classroom. The other 14 grade 3 teachers completed the in-service training and implemented Magic Circle in their classrooms. The control group consisted of all grade 3 students and their teachers in the seven control schools. There were 14 classrooms in each condition: seven pure grade 3 classes and seven mixed classes.

At the beginning of the study, 251 third graders were enrolled in the experimental classes and 281 were enrolled in the control classes. Thirty-four students were excluded from the experimental condition and 31 from the control condition due to lack of parental permission for testing or because they transferred out of the school

during the year. The experimental group consisted of 102 boys and 115 girls, and the control group consisted of 131 boys and 119 girls. The ethnic composition of the sample was 86% (N = 403) White and 14% (N = 64) minority, with Mexican-Americans comprising the largest single minority group (N =29).

In-Service Training Program

Ten weekly two-hour sessions of Magic Circle in-service training were conducted by a highly experienced teacher trainer. The curriculum included both the basic skills for conducting Magic Circles and supplementary skills for improving the quality of teach-student interactions within the Circles and in other classroom activities. The 14 experimental teachers who attended the training sessions and implemented Magic Circle in their classrooms were paid a $200 stipend and were offered graduate-level credits through a local university.

In the first several training sessions, the trainer introduced the teachers to the basic skills necessary for conducting Magic Circle discussions. The trainer showed a videotape of a Magic Circle, led Magic Circles in which the teachers participated, and gave each teacher the opportunity to lead practice Circles. Beginning with the second session, the trainer conducted classroom activity reviews in which the teachers discussed their experiences applying the skills and activities during the prior week.

The next several sessions stressed skill development in three areas: leading the Circle, enforcing Circle rules, and dealing with

children who share inappropriate or personal information in the Circle. Two types of Magic Circle discussion topics were introduced: "awareness" and "mastery" topics. At the beginning of each session, the trainer introduced a self-concept enhancement activity which the teachers could implement in their classrooms.

The last several sessions focused on dealing with special problems that individual students present (e.g., the shy child, the child who answers irrelevantly). In these sessions, the trainer provided instruction in listening and other communication skills. One of the communication skills was "reflecting feelings," a technique designed to identify for speakers the feeling they have conveyed. Teachers were introduced to "social interaction" discussion topics; they learned how to generate useful ideas via "brainstorming"; and they learned how to give specific or "tailored" feedback to students.

During the school year, the trainer visited each of the participant teachers' classrooms a minimum of four times. During some of the initial visits, the trainer led Magic Circles or demonstrated other skills for the teacher. During most of the visits, the trainer observed the teacher conducting a Circle and later provided feedback. On these occasions, immediately after leaving the classroom, the trainer completed a process evaluation form based upon notes taken during the observation.

Instrumentation and Measures

Process evaluation data. Several methods were used to monitor implementation of Magic Circle, and to ascertain reactions to the Circle

training. These methods included a) detailed records of all in-service teacher training sessions, prepared by an expert "outside" observer, b) records of teacher attendance at training sessions, c) surveying of teachers after each training session to obtain feedback on the sessions, d) individual interviews with selected teachers and principals to obtain more detailed information on reactions to the training, e) questionnaire surveys of teachers after training and again at the end of the school year to monitor implementation and to assess the adequacy of the training, f) classroom observations by the in-service trainer of the teachers while conducting Circles with their students, g) weekly reports from teachers on how often they conducted Circles in their classrooms, and h) interviews with selected control school teachers and ineligible experimental school teachers to identify any diffusion of the training. These methods and the process evaluation data gathered with them are described in the next major section of the paper.

Student self-report outcome data. Student pretest and post-test self-report data were obtained with the Self Observation Scales (Intermediate Level, Form C) developed by Stenner and Katzenmeyer (Note 4). This nationally normed instrument has empirically determined scales which measure the ways children perceive themselves and their relationships to their peers, their teacher, and their school. The questionnaire contains 60 statements to which students respond "yes" or "no."

On the post-test, data were obtained from students with the Student Questionnaire (ELE-2), an instrument developed for this study. This instrument consists of two sections. Part 1 contains 18 items selected from the Intellectual Achievement Responsibility Questionnaire developed by Crandall, Katkovsky, and Crandall (1965). Selection of items was based upon published psychometric properties. Four items referring to "parents" were adapted to read "an adult who knows you" in order to conform to California Education Code. These items measure the belief in one's own control over, and responsibility for, intellectual-academic successes and failures.

Part 2 of the Student Questionnaire contains three sets of items: a) the Scholastic scale from the Intermediate-Level of the Self Appraisal Inventory (20 items) developed by the Instructional Objectives Exchange (1972b), a criterion-referenced measure of academic self-esteem; b) the Authority and Control (12 items) and Interpersonal Relationships With Pupils (12 items) scales from the Intermediate-Level of the School Sentiment Index developed by the Instructional Objectives Exchange (1972a), a criterion-referenced measure of attitudes toward school; and d) a measure of perceived peer attitudes toward school developed for this study by adapting 11 items from eight instruments that measure attitudes toward school.

The measures employed in the data analysis for the present study were derived from a theoretical model and empirical scaling analyses. The details of the scaling procedures and results have been

reported by Moskowitz, Condon, Brewer, Schaps, and Malvin (Note 5). Pretest and post-test data were collected from large samples of students in grades 4-6 in addition to the third grade students in this study. Random samples were selected from each grade level. Item means, variances, and intercorrelations were found to be similar across grade levels; thus, scaling procedures were applied across grades. Item intercorrelations from the pretest and from the post-test data were subjected to separate multiple group confirmatory factor analyses.[2] The resultant pretest and post-test scales appear in Table 2 with the number of items contained in each, and their internal consistency reliabilities estimated by coefficient alpha. The final scales included pretest and post-test measures of social self-esteem, attitudes toward peers, affective teaching climate, and attitudes toward school. Other post-test measures included academic self-esteem, locus of control for success, locus of control for failure, and perceived peer attitudes toward school.

Student archival outcome data. In addition to the self-report data gathered directly from students, we obtained achievement data collected by the school district as part of its regular testing activities. For our pretest measure (Prior Read) we employed reading comprehension stanine scores from the Stanford Diagnostic Reading Test (Level Green) administered in September, 1978. For the 10% of students who missed this test, we substituted total reading stanine scores from the Stanford Achievement Test (Primary Level II) administered the prior May.

---

[2]The matrices were computed using pair-wise deletion of missing data and communalities were inserted into their diagonal elements.

TABLE 2

STUDENT PRETEST AND POST-TEST SCALES, NUMBER OF ITEMS AND INTERNAL CONSISTENCY RELIABILITIES (COEFFICIENT ALPHA)

	PRETEST[a]		POST-TEST[b]	
Subscale	Number of Items	Reliability	Number of Items	Reliability
Social Self-Esteem (Social Self)	6	.62	6	.66
Attitudes Toward Peers (Att Peers)	8	.76	8	.80
Affective Teaching Climate (Affec Climate)	7	.74	12	.91
Academic Self-Esteem (Acad Self)	NA[c]	--	11	.79
Locus of Control: Success (Control Suc)	NA	--	7	.56
Locus of Control: Failure (Control Fail)	NA	--	7	.62
Attitudes Toward School (Att School)	6	.68	6	.74
Perceived Peer Attitudes Toward School (Peer Att Sch)	NA	--	8	.72

[a] N = 517

[b] N = 513

[c] NA indicates that this scale was not administered.

According to the test publisher (Karlsen, Madden, and Gardner, 1976), these two tests correlate about .87. For our post-test measures (Read Math), we employed the total reading and total mathematics raw scores from the Stanford Achievement Test, Primary Level III (Madden, Gardner, Rudman, Karlsen, and Merwin, 1973), administered in May, 1979.

Two different measures of student attendance were taken. The total number of unexcused absences for the second semester (Unex Abs) was obtained for each student from the school district's records. This type of absence occurs when a student is absent from school and does not provide the school with a parental excuse indicating that the student was sick. In addition, the average monthly number of absences for each student (Total Abs) during the fifth through eighth school months (January through April) was extracted from teachers' records by project staff.

Students' sex and ethnicity were determined from school district records.

Teacher outcome data. Teacher pretest and post-test self-report data were obtained with different versions of the Teacher Questionnaire, an instrument developed for this study. This instrument included measures of a) teacher satisfaction adapted from the Purdue Teacher Morale Inventory (Rempel and Bentley, 1964); b) faculty cohesiveness adapted from the Teacher Cooperation scale of the Teacher Attitude and Classroom Climate Questionnaire (Kaufman, Semmel and Agard, Note 6) and from the Intimacy scale of the Organization Climate Description Questionnaire (Halpin and Croft, 1963);

and c) the importance and effectiveness of achieving teaching objectives related to Magic Circle, a measure developed for this study. The post-test also included a measure of Magic Circle skill implementation. This measure was developed for the study and consisted of an inventory of classroom behaviors relevant to the Magic Circle in-service training. The number of items included in the pretest and post-test scales and their internal consistency reliabilities estimated by coefficient alpha are shown in Table 3 for grade 3-6 teachers.

Teacher ratings of student misbehavior were obtained at the post-test with the Student Behavior Report. This instrument contained the class roster for each teacher and asked the teacher to indicate how frequently each child had been minor (Minor) and major (Major) discipline problems during the past four months (January through April) on a five-point scale ranging from "never" to "about once a day or more."

Data Collection Procedures

Student survey. The student pretest was administered in October, 1978, and the post-test in May, 1979, by six carefully trained substitute teachers. The Self Observation Scales were administered at both times. In addition, the Student Questionnaire was administered at the post-test. The questionnaires were administered in a single session in the students' classrooms. For the pretest, the items were administered orally by the administrators. For the post-test, the students read the items to themselves (assisted by the administrator when necessary). Two make up sessions were held for students who were absent from the original session.

TABLE 3

TEACHER PRETEST AND POST-TEST SCALES, NUMBER OF ITEMS, AND INTERNAL CONSISTENCY RELIABILITIES (COEFFICIENT ALPHA)

	PRETEST[a]		POST-TEST[b]	
Subscale	Number of Items	Reliability	Number of Items	Reliability
Magic Circle Objectives (MC Object)	8	.80	8	.78
Teacher Satisfaction (Teacher Satis)	3	.68	8	.80
Faculty Cohesiveness (Faculty Cohes)	7	.87	7	.90
Magic Circle Skill Implementation (MC Skill)	NA[c]	--	16	.84[d]

[a] N = 73

[b] N = 60

[c] NA indicates that this scale was not administered

[d] N = 26

Teacher survey. The teacher pretest was administered in September, 1978, during meetings at each school conducted by project staff. Teachers completed the post-test questionnaires in their classrooms while their students were being surveyed. To ensure the privacy of their responses, teachers were provided with questionnaires containing unique identifiers.

Process Evaluation Methods and Findings

Descriptions of In-Service Training Sessions

To document the in-service training, a skilled observer attended all training sessions and made detailed notes about both the trainer's and the participating teachers' activities. The observer later prepared lengthy descriptions of a) the agenda as planned and as actually followed, b) any materials used or distributed, c) the trainer's presentations, and d) the trainer's and teachers' verbal and nonverbal interactions.

The observer sat in the back of the room and avoided interactions with the teachers. The trainer explained that the observer was present to "take notes so that we can develop a training manual."

These observations showed that the trainer was able to gain the trust and confidence of the teachers, who talked frankly about their problems and successes in implementing Magic Circle techniques. The teachers seemed to enjoy learning the Circle skills, to take the training seriously and to participate actively, and to appreciate the emotional support and encouragement that the group provided.

Teacher Attendance at Training Sessions

The teachers were required to attend nine of the ten in-service sessions in order to receive a $200 stipend. One of the 15 particjpating teachers dropped out after the third session because she was opposed to Magic Circle's goals and methods. All of the other 14 teachers completed

the training and attended nine or more sessions. Six of these teachers also attended a "reunion" session held two months after the end of training.

Principals from the experimental schools were also invited to attend the training, and all of the six eligible principals attended one or more (up to six) sessions.

Teacher Feedback on the Individual Training Sessions

At the end of each training session, teachers were asked to complete a one-page "feedback form" anonymously. This form solicited ratings of, and comments about, the sessions. On the form, teachers rated each session for interest, organization, usefulness, and enjoyableness, using five-point scales with higher numbers signifying more positive ratings. Ratings for the individual sessions all averaged 4.25 or above on each of the four measures, and 26 of the 40 average session ratings were at or above 4.70, indicating that the training was very favorably regarded by the teachers. This high regard was also apparent from the teachers' very complimentary written comments, and from the almost complete absence of critical comments.

Interviews with Participating Teachers and Principals

Three experimental school principals and 4 of the 14 participating teachers were interviewed individually by a project researcher in April, 1979, approximately three months after the training had ended. Principals and teachers were randomly selected for the interviews, which were semi-structured and open-ended. The interviews were tape-recorded with the respondent's permission, and lasted 15-30 minutes.

In the interviews, all of the principals and teachers praised the training extensively. All felt the training was useful, enjoyable, well organized, and relevant to their needs. Three of four teachers thought that the trainer's classroom follow-up visits were very helpful and supportive; the fourth said the visits were pleasant but unnecessary because she was having no difficulty implementing Magic Circle skills. All of the teachers said that the training fully prepared them to use Magic Circle effectively.

The interviews revealed that certain in-service skills were used less frequently than others, most notably, a) brainstorming techniques and b) teaching students to lead Circle discussions. They also indicated that the training may have been too lengthy; several principals and teachers thought that the last few sessions were redundant and unnecessary.

All four teachers reported that their students thoroughly enjoyed participating in Circles, and looked forward to them. All reported positive effects on students, both in and out of Circles, including increased cooperation and respect, less teasing and criticizing, heightened awareness of others' feelings, and increased feelings of community in the classroom. All of the teachers planned to use Magic Circle the following year.

Questionnaire Surveys of Participating Teachers

At the completion of training (January, 1979), and again near the end of the school year (May, 1979), participating teachers provided detailed information regarding implementation by completing lengthy questionnaires. On the questionnaires, the teachers reported how often

they used each in-service skill, rated the usefulness of each skill, and rated their own mastery of each skill.

Table 4 shows year-end data on frequency of implementation, and both mid-year and year-end data on usefulness and mastery of the skills. The first five skills listed in Table 4 are used only when conducting Circles, the second set of five skills can be used in Circles or in general teaching, and the last "skill"--self-concept enhancing activities--is a series of classroom exercises used to supplement the basic Circle activity.

The data regarding frequency of implementation in Table 4 indicated that at year-end, the teachers were using nearly all of the skills at least several times per month, and many of the skills at least several times per week. However, two skills were not used this frequently: preparing students to lead Magic Circles, and brainstorming. This implementation pattern is consistent with the interview data reported above, and suggests that the teachers were continuing to use both the Circle-specific skills and the "general" skills several months after the training had been completed.

With regard to the usefulness of the skills, Table 4 shows that with the exceptions of student-led Circles and brainstorming, the teachers rated all of the skills as highly useful at both mid-year and year-end. The teachers' average ratings of their own mastery of the skills clustered in the area between "good" and "excellent," although several were a bit lower. Mastery ratings at year-end were generally similar to those at mid-year.

TABLE 4

TEACHERS' REPORTS ON FREQUENCY, UTILITY, AND QUALITY OF SKILL IMPLEMENTATION AT MID-YEAR AND END OF YEAR

Classroom Teaching Skill	At End of School Year, Percentage of Teachers Using the Skill at Least Several Times Per:		Mean Ratings of Value of the Skill (1=High Utility; 5=Low Utility)		Mean Ratings of Mastery of Skill[a] (1=Excellent; 4=Poor)	
	Month	Week	Mid-Year	End-Year	Mid-Year	End-Year
Introducing, explaining and restating the topic	92 %	58 %	1.15	1.17	1.69	1.67
Enforcing the four rules in conducting the discussion	92	58	1.30	1.17	1.46	1.58
Conducting the cognition stage and ending the circle	92	50	1.23	1.17	1.92	2.08
Dealing with shy, copying, irrelevant answering children	92	50	1.33	1.75	2.46	2.33
Preparing students to lead Magic Circle	33	0	3.36	3.00	3.25	2.00
Leading brainstorming activities	50	17	2.18	2.55	2.29	1.88
Dealing with disruptive students	90	70	1.46	1.42	1.92	1.67
Using I-messages	100	75	1.31	1.33	1.46	1.83
Using selected parroting	100	83	1.92	1.42	1.85	1.73
Using descriptive, non-judgmental terminology	100	91	1.31	1.25	2.15	2.00
Using self-concept enhancing activities	100	67	1.23	1.27	2.15	2.00

[a] These ratings completed only by teachers who have used the skill in their classroom.

To gather additional feedback on the in-service training, the mid-year questionnaire also asked the teachers to evaluate the in-service course as a whole, using five-point rating scales. These ratings were very favorable with respect to interest ($\underline{M}$ = 4.84), organization ($\underline{M}$ = 4.92), usefulness ($\underline{M}$ = 4.69), and enjoyableness ($\underline{M}$ = 4.92). The trainer was also highly rated along a number of dimensions. On the year-end questionnaire, the trainer's classroom follow-up visits received a 4.36 average rating for usefulness.

To summarize, data from the two questionnaire surveys indicate that teachers were using most of the skills on a regular basis, were finding the skills useful, considered themselves adept at using the skills, and positively evaluated the training and the trainer.

Trainer Observations of Classroom Implementation

The Magic Circle trainer visited each classroom several times during the school year to observe the teachers leading Magic Circles. After each classroom observation, the trainer reported a) the number of times each skill was used, b) the quality of implementation of each skill, and c) the number of occasions on which each skill could have been used appropriately but wasn't (called "missed occasions").

Averaged across visits to each classroom, the observational data regarding quality of implementation indicate that all 14 teachers were using the basic Magic Circle skills with at least moderate proficiency.

With respect to these basic skills, the trainer's ratings of the teachers' quality of implementation ranged between 3.34 and 5.00 on a five-point scale, with 11 of the 14 teachers receiving ratings of 4.00 or better. The average rating was 4.43 (SD = .45).

Across all in-service skills (i.e., both the specific and general skills), the average quality rating given each teacher by the trainer ranged between 3.84 and 4.88 (M = 4.41, SD = .26). Thus, in the trainer's estimation, all 14 teachers were also implementing the general skills with good or excellent proficiency.

Several of the skills were only infrequently observed by the trainer. These were a) preparing students to lead Circles, b) dealing with shy-copying-irrelevant answering children, and c) brainstorming. No more than four teachers were observed using any of these skills, and no teacher was observed using these skills more than an average of once per visit. In contrast, at least nine teachers were observed using each of the other skills, and with substantially higher frequencies.

For most of the skills, the trainer noted few missed occasions relative to the observed frequencies of implementation. Thus, in the trainer's estimation, the teachers were applying the skills in most of the appropriate instances.

In summary, data obtained from the trainer's classroom observations confirmed that the teachers were using most (but not all) of the in-service skills, that they were using the skills proficiently, and that they were taking advantage of most opportunities to apply the skills appropriately.

Classroom Implementation Process Data

Beginning with the onset of training, and continuing through the school year, experimental teachers completed a weekly log indicating a) how many Circle sessions were held during the week, b) which students attended each session, and c) the reasons for any absences. Twelve of the fourteen teachers consistently completed and returned the weekly log. An average of 23.4 weeks of data (SD = 1.4) were collected from these 12 teachers. The other two teachers returned only 7 and 13 forms respectively, ignoring many reminders.

Some of the teachers deviated from the instructions in completing the logs at various times during the year. This source of error made the data questionable at the student level; hence a weekly class-level index of exposure to Magic Circle was constructed. By summing this weekly index over the number of weeks of data collected from each teacher, a conservative estimate was obtained of the total number of Circles that the typical student in each class received during the year (Circle Quantity). This estimate ranged from 8 to 36 sessions, and indicated that students in five classes received less than 18 Circles over the school year. The average class of students participated in 22 sessions (SD = 8.7) which amounts to almost one Circle (.87) per week over 25 weeks.

Interviews With Control School Teachers and Ineligible Experimental School Teachers

To check for any diffusion of the treatment, interviews were conducted in April, 1979 with four principals and three teachers at

control schools, and with one nonparticipating principal and three teachers at the experimental schools. The interviews were conducted by a member of the research team, and were semi-structured and open-ended.

Only one of the control school principals and none of the control school teachers had more than a vague understanding of Magic Circle's purpose and methods. The one principal who could describe the strategy had recently learned about it in a discussion with the Magic Circle trainer. Among the nonparticipating principal and teachers from the experimental schools, only the principal had any substantive knowledge of the strategy, and he had learned what he knew from attending the second in-service training session.

In summary, when interviewed near the end of the school year, the control and nonparticipating teachers and administrators knew little or nothing about Magic Circle's goals or techniques, so that diffusion of the treatment was not a plausible threat to the internal validity of the evaluation design.

Results

A bi-level analysis approach was adopted for several reasons. The experimental design employed in this study involved randomly assigning schools (and not classes or students) to the experimental or control condition. Least-squares analysis techniques assume the statistical independence of observations. Treating individual students' responses as independent when they share a common school and classroom environment (and common exposure to treatment) is problematic because the likelihood of spuriously obtaining a significant treatment effect is enhanced. In addition, student data and classroom aggregate data sometimes possess different substantive meaning and may be subject to different processes resulting in different findings. Thus, to complement the student-level analyses, we also conducted class-level analyses of the student data. We chose to examine class-level rather than school-level data because student interdependence occurs primarily within classes and not within schools. In addition, teachers were responsible for the implementation of the treatment, and they are assigned to particular classrooms.

Treatment of Missing Data

The problem of missing observations arose at two points in this analysis. In each instance, we adopted a strategy that allowed maximum use of the available data points while minimizing outcome bias.

Computation of Scale Scores. A scale score was computed for a student if at least 60% of the items comprising that scale were present. Any missing item score was replaced by the mean for that item in the appropriate subclass of the experimental design. This procedure utilized most of the item data and provided unbiased cell means. However, it constrained cell variances and inflated degrees of freedom artificially. When less than 40% of the items comprising a scale were missing, the student received a missing value for that scale.

Missing Scale Scores in Multivariate Analyses. Each of the multivariate analyses in this report was computed from summary data (i.e., subclass means and frequencies and a pooled within-class variance-covariance matrix). Choice of this input option enabled better utilization of data than typically is possible with available statistical routines, which exclude a case with even a single missing datum from analysis (c.f., MULTIVARIANCE: Finn, 1976; SPSS: Cohen and Burns, Note 7). Specifically, the subclass mean and variance for a variable were based on all cases for which the variable was present, and the computation of each subclass covariance included every case in which the two variables were present. Hence, the number of observations contributing to covariances differed slightly within each subclass matrix. To compensate for our earlier inflation of degrees of freedom, we chose the minimum pairwise frequency to represent each subclass both in computing the pooled matrix and later for significance testing.

Analysis of Initial Equivalence and Attrition

An analysis was conducted to determine a) whether students in the experimental and control conditions were equivalent on the pre-test measures and b) whether attrition had a differential effect on these conditions. Initial nonequivalence between treatment conditions has implications for both internal validity (the ability to make causal inferences about the effect of the treatment) and external validity (the ability to generalize findings to other populations). Attrition may affect internal validity if students missing from one condition differ systematically in kind from those missing from another condition. Attrition may also affect external validity if the missing (attrited) students differ systematically from those who remain in the sample (non-attrited).

As suggested by Jurs and Glass (1971), a two-way multivariate analysis of variance (MANOVA) was performed on the student pretest data. The factors of the design were treatment condition (experimental versus control), and attrition status (attrited versus non-attrited). A main effect for treatment condition would provide evidence for initial non-equivalence. A main effect for attrition status would provide evidence for limited external validity. The interaction between treatment condition and attrition status would indicate a threat to internal validity.

None of the tested effects was significant. The respective F-ratios for treatment condition, attrition status, and the interaction term

were 0.91, 1.60, and 1.56, with 5 and 414 degrees of freedom. Hence, the experimental and control students were initially equivalent, and attrition did not impair this equivalence.

Multivariate Analyses of Student Data

Some prior research has indicated that the effect of Magic Circle may be a function of the students' prior achievement level. A preliminary analysis was performed to explore this hypothesis by subjecting the post-test data to MANOVA employing treatment condition (T), sex of student (S), and prior reading level (P), as factors in the design. Prior reading level was used as an indicator of prior achievement. This variable was obtained by trichotomizing Prior Read. The results of this analysis indicated that neither the T x P interaction, $F(28,690) = 1.18$, nor the T x S x P interaction, $F(28,690) = 0.99$, was significant. Thus, no support was found for the hypothesis that students' prior reading level was related to the effects of Magic Circle. Hence, the design employed in subsequent analyses includes only treatment condition and sex (of student) as factors.

For the multivariate analysis of the student data, a complete factorial design was employed for hypothesis testing, with treatment condition and sex as the factors. Multivariate analysis of covariance (MANCOVA) was performed using pretest measures as covariates when the appropriate assumptions were met; otherwise, MANOVA was performed on post-test outcomes. Since the design was nonorthogonal due to unequal cell sizes, the reported main effect for each factor was tested after the variance associated with the other factor had

been removed and the treatment x sex interaction was tested after removing the main effect variance. Effects were estimated subsequently using a reduced-rank model based upon the prior hypothesis testing. When significant interactions were obtained, the simple treatment effects were tested by employing the Bonferroni t statistic (Myers, 1979). This statistic is a t-test that controls the Type I error rate for the number of tests performed. Because of space limitations, only results pertaining to the treatment condition and treatment interaction are discussed.

The post-test measures were divided into four subsets based upon *a priori* hypotheses about their interrelationships and empirical findings concerning the homogeneity of their covariance structures. One subset (A) consisted of attitudes toward one's social self and social relationships: social self-esteem (Social Self) and attitudes toward peers (Att Peers).

A second subset (B) consisted of attitudes and behaviors related to one's academic self: locus of control: success (Control Succ), locus of control: failure (Control Fail); affective teaching climate (Affec Climate); academic self-esteem (Acad Self); attitudes toward school (Att School); perceived peer attitudes toward school (Peer Att Sch); reading achievement (Read); and mathematics achievement (Math). A third subset (C) included the two attendance measures, unexcused absences (Unexc Abs) and total absences (Total Abs), and the final subset (D) contained the teacher ratings of student minor (Minor) and major (Major) misbehavior.

Each of the four subsets of post-test measures was analyzed separately. Preliminary analyses were performed to identify the set of pretest outcomes that best predicted each set of post-test outcomes and did not violate the assumptions of the covariance analysis (primarily, that the regression hyperplanes be homogeneous across the cells of the design). The final sets of covariates selected for the analyses included the following pretest measures for set A: Social Self and Att Peers; and for set B: Social Self, Att Peers, Affec Climate, School Att, and Prior Read. For sets C and D, covariate selection was abandoned because either the pretest measures were not good predictors of the post-test outcomes or the homogeneity assumptions were violated. For each of these sets of measures the MANOVA results have been reported.

Results of the multivariate analyses of the student data have been summarized in Table 5. For Set A, a marginally significant treatment main effect, $F(2,384) = 2.80$, $p < .07$, and treatment x sex interaction, $F(2,384) = 2.84$, $p < .06$, were obtained. For Set D, a highly significant treatment main effect, $F(2,411) = 19.46$, $p < .0001$, and treatment x sex interaction, $F(2,411) = 8.54$, $p < .0003$, were found. No treatment-related effects were significant in Sets B and C.

The only reliable univariate effects on Set A measures occurred for the treatment main effect, $F(1,385) = 5.23$, $p < .03$,

TABLE 5

SUMMARY OF MULTIVARIATE ANALYSES OF STUDENT DATA

Source of Variation	Subset of Measures	df	Multivariate F	p<
Treatment Condition (T)	A[a]	2,384	2.80	.07
	B[b]	8,371	1.40	ns
	C[c]	2,458	0.04	ns
	D[d]	2,411	19.46	.0001
Sex (S)	A	2,384	1.90	ns
	B	8,371	3.15	.002
	C	2,458	1.42	ns
	D	2,411	24.55	.0001
T x S	A	2,384	2.84	.06
	B	8,371	1.04	ns
	C	2,458	0.59	ns
	D	2,411	8.54	.0003

[a]Subset A analysis: Variates and covariates were Social Self and Att Peers.

[b]Subset B analysis: Variates were Control Succ, Control Fail, Affec Climate, Acad Self, School Att, Peer Att Sch, Read, and Math. Covariates were Social Self, Att Peers, Affec Climate, School Att, and Prior Read.

[c]Subset C analysis: Variates were Unexc Abs and Total Abs.

[d]Subset D analysis: Variates were Minor and Major.

and the treatment x sex interaction, $F(1,385) = 5.17$, $p < .03$, on Social Self. Examination of the simple treatment effects by sex yielded a significant effect for boys, $t(385) = 3.12$, $p < .01$, but not for girls, $t(385) = 0.03$. Inspection of the means revealed that experimental boys scored .45 standard deviations higher than control boys on the pretest-adjusted measure of social self-esteem.

From Set D, reliable univariate effects were found on both measures, Minor and Major. The respective treatment main effects were $F(1,412) = 16.45$, $p < .0001$, and $F(1,412) = 6.11$, $p < .02$. The Treatment x Sex interactions were $F(1,412) = 4.28$, $p < .04$, and $F(1,412) = 5.30$, $p < .03$, respectively. Examination of the simple treatment effects by sex revealed no reliable differences for girls on Minor, $t\ (412) = 1.51$, or on Major, $t\ (412) = 0.20$. For boys, however, a negative treatment effect was obtained on Minor, $t\ (412) = 4.30$, $p < .002$, and a positive effect on Major, $t\ (412) = 3.38$, $p < .002$. Experimental boys were generally rated 0.61 standard deviations higher than control boys on the minor misbehavior measure and 0.48 standard deviations lower than control boys on the major misbehavior measure.

Analyses of Class-Level Student Data

For the class-level student analysis each class was divided into two analysis units, one consisting of the boys in the class and one

consisting of the girls. For each such unit, mean scores were computed on the student measures, and separate analyses were conducted for the boys and girls. With this approach, units of analysis are more interdependent than in a strictly class-level analysis, but the approach controls for, and enables examination of, the influence of student sex upon the treatment effects.

Univariate analyses of variance were applied to each pretest measure with treatment condition as the independent variable. The results indicated no initial differences between the treatment and control groups for either the boys or the girls. Since no initial differences were obtained, a one-way treatment condition univariate analysis of variance was performed on each post-test measure. The results of these analyses are summarized in Table 6. For the boys the class-level results were consistent with the student-level results. Significant treatment effects were obtained on two measures, Social Self, $F(1,26) = 16.19$, $p < .001$, and Minor, $F(1,24) = 5.16$, $p < .04$, and a marginally significant effect was obtained on a third, Major, $F(1,24) = 3.07$, $p < .09$. Congruent with the student-level findings, positive treatment effects were obtained on Social Self and Major, and a negative treatment effect on Minor.

For the girls the class-level results were consistent with the student-level results as no significant treatment effects were obtained.

Analyses of Teacher Data

A one-way treatment condition univariate analysis of covariance was applied to eac post-test teacher measure with the corresponding

TABLE 6

SUMMARY OF UNIVARIATE ANALYSES OF CLASS-LEVEL STUDENT DATA

Measure	Male Classes			Female Classes		
	df	F	p <	df	F	p <
Social Self	1,26	16.19	.001	1,26	0.02	ns
Att Peers	1,26	1.76	ns	1,26	0.30	ns
Affec Climate	1,26	0.00	ns	1,26	2.13	ns
Acad Self	1,26	0.83	ns	1,26	1.73	ns
Control Succ	1,26	0.03	ns	1,26	0.58	ns
Control Fail	1,26	0.55	ns	1,26	0.14	ns
Att Sch	1,26	1.07	ns	1,26	0.15	ns
Peer Att Sch	1,26	0.01	ns	1,26	0.87	ns
Read	1,26	0.35	ns	1,26	0.04	ns
Math	1,26	1.14	ns	1,26	0.06	ns
Unexc Abs	1,26	0.43	ns	1,26	0.20	ns
Total Abs	1,26	0.15	ns	1,26	0.73	ns
Minor	1,24	5.16	.04	1,24	2.26	ns
Major	1,24	3.07	.09	1,24	0.05	ns

pretest used as a covariate. Since the corresponding pretest did not exist for MC Skill, analysis of variance was performed on this measure.

Results of the univariate analyses of variance and covariance on the teacher data are summarized in Table 7. Significant treatment effects were obtained on two of the four teacher measures, Teacher Sati., $F(1,21) = 10.13$, $p < .005$ and MC Skill, $F(1,24) = 11.50$, $p < .003$. Experimental teachers generally rated themselves as more satisfied with teaching on the post-test (adjusted for their pretest scores) than did control teachers. Experimental teachers also scored higher than controls on the behavioral inventory reflecting classroom implementation of Magic Circle-related skills.

The Relationship of Student Outcomes to Treatment Implementation

The analyses reported thus far involved comparisons of treated versus untreated students, teachers, and classrooms. Just as these groups may differ on the basis of presence or absence of the Magic Circle treatment, the experimental classes might also differ as a function of the "amount" of treatment received. Therefore, exploratory analyses were conducted to investigate the relationship among the experimental classrooms' average student post-test outcomes and the classrooms' amount of exposure to Magic Circle. Classroom exposure to Magic Circle (Circle Quantity) was measured by an index constructed from the teachers' weekly reports described earlier in the section on process evaluation. This measure was dichotomized at the median resulting in two groups of experimental classrooms, a high implementation group and a low implementation group.[3] The high implementation group received

[3] While we recognize that a median split sacrifices possibly meaningful data, we were unwilling to assume a linear relationship between implementation and student outcomes, an assumption necessary for a correlational analysis approach.

TABLE 7

SUMMARY OF UNIVARIATE ANALYSES OF TEACHER DATA

Measure	Analysis Type [a]	Treatment Main Effect F	df	p <
MC Object	C	0.05	1,21	ns
Teacher Satis	C	10.13	1,21	.005
Faculty Cohes	C	0.92	1,21	ns
MC Skill	V	11.50	1,24	.003

[a] C = analysis of covariance
V = analysis of variance

an average of 28.7 Circles (SD = 3.8), and the low implementation group averaged 14.9 Circles (SD = 6.3) over the school year.

Analyses of variance were performed on the class mean student pretest data to determine whether high and low implementation classes initially differed on these measures. The results indicated no initial differences for the boys. The girls differed on one pretest measure, Prior Read, $F(1,12) = 3.84$, $p < .08$, with the high implementation group initially having higher reading scores.

For the boys, since no initial differences were obtained, univariate analyses of variance were performed on the class mean student post-test data with implementation level (high versus low implementation) as the independent variable. Results from these analyses are summarized in Table 8. No significant differences were obtained; that is, the level of treatment implementation did not have a detectable effect upon any student measure. Of course, the statistical power of these analyses was limited due to the small number of classes that were available ($N = 14$).

For girls, univariate analyses of covariance were performed on the class mean student post-test data with implementation level as the independent variable and Prior Read as the covariate. These analyses partially control for the initial differences between the groups. On six of the measures, analyses of variance of the post-test data were reported as the assumptions of the covariance analyses were not met. The results have been summarized in Table 8. A significant difference was obtained on one of the fourteen measures, Unexc Abs, $F(1,11) = 5.34$,

TABLE 8

SUMMARY OF TREATMENT IMPLEMENTATION ANALYSES

Measure	Male Classes F(1,12)	p <	Female Classes F(1,11)	p <
Social Self	0.72	ns	1.46[a]	ns
Att Peers	0.02	ns	0.07	ns
Affec Climate	0.74	ns	1.75[a]	ns
Acad Self	0.28	ns	0.55	ns
Control Succ	0.21	ns	0.00	ns
Control Fail	0.76	ns	0.05	ns
Att Sch	0.28	ns	0.14[a]	ns
Peer Att Sch	3.02	ns	1.87	ns
Read	0.71	ns	0.30[a]	ns
Math	2.81	ns	0.04	ns
Unexc Abs	0.66	ns	5.34	.05
Total Abs	3.12	ns	0.01[a]	ns
Minor	0.00	ns	0.06	ns
Major	0.62	ns	1.22[a]	ns

Note: The independent variable is treatment implementation level (high versus low). For the males, analyses of variance of the class mean data have been reported For the females, analyses of covariance of the class mean data using Prior Read as a covariate have been reported (see superscript).

[a]Analysis of variance was performed, F(1,12), due to violations of covariance assumptions.

$p < .05$. Girls in the high implementation group had fewer unexcused absences than those in the low implementation group. Given the total number of significance tests performed in these analyses, it is likely that this single observed difference may be a Type I error.

Discussion

The process data indicated that teachers in the experimental group found the Magic Circle in-service training to be interesting, well organized, and useful. Furthermore, the teachers reported that they had mastered most of the skills taught in the training and had implemented them regularly in their classrooms. The only weakness of the training reported by a few teachers was that they found the last few training sessions to be redundant.

The frequency with which teachers conducted Circles in their classroom varied considerably. Over the school year, some classes received one Circle every three weeks, whereas other classes received at least one Circle per week. In spite of this variation in frequency of implementation, the quality of the teachers' Circle leadership, as rated by the trainer, was uniformly high. This observational finding coincided with the teachers' self-reported mastery of the skills.

Consistent with the process results, on post-testing, experimental teachers reported using Circle-related skills in their classrooms to a greater extent that did control teachers. Experimental teachers also reported greater personal satisfaction with teaching than did control teachers (after adjusting for pretest levels of satisfaction). Thus, teachers who participated in the training seem to have benefited from the experience.

As compared with controls, boys in the experimental group generally were rated by their teachers to be major discipline problems to a lesser extent, but minor discipline problems to a greater extent. This finding does not coincide with the teacher interview data since experimental teachers reported only positive effects of Magic Circle upon student classroom behavior. This pattern of results suggests that experimental teachers may have used the two discipline rating scales in a different manner than did the control teachers. The experimental teachers may have employed different definitions of minor and major discipline problems.[4] Such a cognitive bias in rating could possibly be attributed to the Magic Circle training or the experience of leading Circles in the classroom. However, the bias is just as likely to be motivational as cognitive, and may be attributable to self-enhancing, self-defensive or positivity biases, or due to social desirability or compensatory rivalry (Cook and Campbell, 1979). As the validity of these teacher ratings of student misbehavior is questionable, so is the pattern of results obtained from them.

As hypothesized, Magic Circle increased boys' social self-esteem relative to controls. This effect was, however, somewhat puzzling since it obtained only for boys, and not for girls, and boys and girls were initially equivalent on this measure. In addition, it was an isolated effect in that boys' attitudes toward peers, which conceptually and empirically are related to their social self-esteem were not

[4]We had anticipated this possibility and included a structured questionnaire with the Student Behavior Report requesting the teachers to classify 17 different hypothetical misbehaviors as minor, major, or not discipline problems. Unfortunately, most of the teachers misinterpreted the instructions, rendering these data useless.

affected by Magic Circle. Furthermore, improved social self-esteem among experimental boys showed no relationship to the measure of treatment implementation, Circle Quantity. Thus, the increased social self-esteem observed among experimental boys cannot be explained by the number of Circles in which they participated. In fact, none of the boys' outcomes was reliably related to implementation. These findings suggest that even minimal participation in Magic Circle may be sufficient to improve boys' social self-esteem. While this is not a satisfying explanation, it is the only hypothesis which the data do not contradict. A number of alternative hypotheses have been explored through secondary analyses of the data. These have not been described in this paper, and none was found to be tenable.

The student outcome data showed no effects of Magic Circle on girls in the experimental group as compared to control group girls. In addition, among the experimental classes, no relationship was found between the level of treatment implementation and post-test outcomes.

Students in the present study may have received considerably fewer Circles than did students in previous studies, although our method of estimating level of implementation was decidedly conservative. By our estimation method, students received an average of 22 Circles, whereas the average in 13 prior studies, as reported in the HDTI review, was 60 Circles. We do not know how implementation data were collected in other studies, and it may be that some of them were biased in a liberal direction. Several of these studies involved very small numbers of students and teachers, and the implementation levels achieved in them

may be atypical. Also, in several studies "outside" personnel rather than classroom teachers were used to conduct Circles. Given the numbers involved in the present study, and that several efforts were made to stimulate voluntary implementation by the teachers, the obtained levels may represent a more realistic approximation of what can be achieved through voluntary involvement of regular classroom teachers.

It is also noteworthy that in prior studies cited by HDTI, no relationship existed between implementation level and student outcomes. Thus, the plausible expectation that greater implementation will produce more positive treatment effects has not been confirmed as yet.

In the present study, although a strong research design was utilized and many types of process and outcome data were collected, the pattern of results was largely unclear. Despite various inconsistencies, three general conclusions seem warranted:

1. Teachers who participated in the Magic Circle in-service training found it interesting and useful. They mastered most of the Magic Circle skills and implemented them in their classroom to varying degrees. Their personal satisfaction with teaching improved more than the control teachers.

2. In comparisons of the experimental and control group, Magic Circle was found to have little detectable positive effect upon the student outcomes examined. The one notable effect was upon boys' social self-esteem.

3. The relationship between the treatment process and its effect upon the experimental boys appears to be complex and may be a function of process variables not measured in the current study (e.g., the specific curriculum employed by the teachers in the Circle sessions).

Resolution of this puzzling pattern of results may come with continued delivery of services to the experimental cohort during the next two years.

Students in the experimental group are scheduled to receive a second year of Magic Circle via their fourth grade teachers, and other interventions via their fifth grade teachers. Thus, if the inconsistent pattern of results is a function of the limited duration and intensity of services delivered to date, the second and third years of this study may show a more coherent and decisive set of effects.

REFERENCES

Cook, T. & Campbell, D. *Quasi-Experimentation: Design and Analysis Issues for Field Settings.* Chicago: Rand McNally, 1979.

Crandall, V., Katkovsky, W. & Crandall, V. Children's belief in their own control of reinforcements in intellectual-academic achievement situations. *Child Development,* 1965, *36,* 91-109.

Finn, J. *MULTIVARIANCE: Univariate and Multivariate Analysis of Variance, Covariance and Regression.* Ann Arbor, Michigan: National Educational Resources, Inc., 1972.

Halpin, A. & Croft, D. *The Organizational Climate of Schools.* Chicago: Midwest Administration Center, University of Chicago, 1963.

Instructional Objectives Exchange. *Attitude Toward School K-12.* (rev. ed.) Los Angeles: Instructional Objectives Exchange, 1972a.

Instructional Objectives Exchange. *Measures of Self-Concept K-12.* (rev. ed.) Los Angeles: Instructional Objectives Exchange, 1972b.

Karlsen, B., Madden, R. & Gardner, E. *Stanford Diagnostic Reading Test: Manual for Administering and Interpreting (Green Level).* New York: Harcourt Brace Jovanovich, 1976.

Madden, R., Gardner, E., Rudman, H., Karlsen, B. & Merwin, J. *Stanford Achievement Tests: Norms Booklet (Primary Level III).* New York: Harcourt Brace Javanovich, 1973.

Myers, J. *Fundamentals of Experimental Design (Third Edition).* Boston: Allyn And Bacon, Inc., 1979.

Rempel, A. & Bentley, R. The measurement of teacher morale: a factor analysis approach. *Educational and Psychological Measurement,* 1964, *24,* 631-643.

REFERENCE NOTES

1. Ball, G. *Magic Circle: An Overview of the Human Development Program.* La Mesa, CA: Human Development Training Institute, Inc., 1974.

2. Ward, K. Human Development Training Institute. Personal Communication, June 27, 1980.

3. Human Development Training Institute. Review of the Literature of Research Investigations of the Magic Circle/Human Development Program, La Mesa, CA: Human Development Training Institute, 1977.

4. Stenner, A. & Katzenmeyer, W. Technical Manual and User's Guide: Self-Observation Scales, NTS Research Corporation, Durham, NC: undated.

5. Moskowitz, J., Condon, J., Brewer, M., Schaps, E. & Malvin, J. The Napa Project: Scaling of Student Self-Report Instruments. Prevention Branch, National Institute on Drug Abuse, 1979.

6. Kaufman, M., Semmel, M. & Agard, J. *Supplemental Materials to Year I Interim Report, Part II.* Washington, D. C.: U. S. Office of Education, Project Prime, undated.

7. Cohen, E. and Burns, P. SPSS-MANOVA-Multivariate Analyses of Variance and Covariance. Northwestern University, Vogelback Computing Center, Document No. 413 (A), 1977.