DOCUMENT RESUME

ED 212 681 TM 820 138

AUTHOR Moskowitz, Joel M.; And Others
TITLE A Process and Outcome Evaluation of a Peer Teaching Primary Prevention Program.
INSTITUTION National Inst. on Drug Abuse (DHHS/PHS), Rockville, Md.; Pacific Inst. for Research and Evaluation, Napa, Calif.
PUB DATE Sep 81
NOTE 45p.; For related documents, see TM 820 136-137.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Drug Abuse; Intermediate Grades; *Intervention; Peer Influence; *Peer Teaching; Program Attitudes; *Program Effectiveness; *Program Evaluation; Student Participation; Student Teacher Relationship

ABSTRACT

This study evaluated the effectiveness of Jigsaw as a strategy to prevent substance abuse. Jigsaw is a structered peer teaching program in which students teach part of the regular classroom curriculum to a group of peers. The goals of Jigsaw were to make student learning more active and student-student interactions more cooperative. As a result of these changes, students were expected to develop more positive attitudes, behaviors and norms regarding self, peers and school. Eventually, these gains should reduce students' acceptance and use of psychoactive substances. The treatment group consisted of 13 fourth-sixth grade teachers and their 261 students. Thirty teachers and their 560 students served as the nonparticipant comparison group. The Jigsaw in-service training consisted of two-hour sessions, held once a week for nine weeks, and one review session held six weeks later. The trainer assisted teachers in their classrooms both before and after the training ended. Participant and nonparticipant 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. (Author/CE)

A Process and Outcome Evaluation of A Peer Teaching Primary Prevention Program

Submitted to:

National Institute on Drug Abuse

Prevention Branch

September, 1981

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

Joel M. Moskowitz, Ph.D.
Janet Malvin, Ph.D.
Gary A. Schaeffer, Ph.D. Cand.
Eric Schaps, Ph.D.
John W. Condon, Ph.D.

ABSTRACT

This study evaluated the effectiveness of Jigsaw as a strategy to prevent substance abuse. Jigsaw is a structured peer teaching program in which students teach part of the regular classroom curriculum to a group of peers. The goals of Jigsaw were to make student learning more active and student-student interactions more cooperative. As a result of these changes, students were expected to develop more positive attitudes, behaviors and norms regarding self, peers and school. Eventually, these gains should reduce students' acceptance and use of psychoactive substances.

The treatment group consisted of 13 4th-6th grade teachers and their 261 students. Thirty teachers and their 560 students served as the non-participant comparison group.

The Jigsaw in-service training consisted of two-hour sessions, held once a week for nine weeks, and one review session held six weeks later. The trainer assisted teachers in their classrooms both before and after the training ended.

Process evaluation data included a) teacher feedback on the individual training sessions, b) questionnaire surveys of participating teachers, c) weekly reports by teachers on implementation, and d) classroom observation. The training was highly regarded by participants. Most of the skills were regularly used by about half the teachers. Teachers found the skills useful and reported their mastery of them as fair to good. There was wide variation in the implementation of Jigsaw activities; teachers reported from .24 to 3.77 hours per week allotted to Jigsaw. The quality of Jigsaw implementation was disappointing. In only four of the 12 classrooms was Jigsaw appropriately implemented. In five classrooms Jigsaw was radically modified by the teachers and in the remaining three classrooms students were largely off-task.

Participant and nonparticipant 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.

There were no significant differences between participant and nonparticipant teachers on three teacher variables. There were several differences between participant and nonparticipant students although they did not replicate across different grade-sex groups. Comparing students in "exemplary" Jigsaw classrooms with nonparticipants yielded four effects that replicated

results from the participant vs. nonparticipant analyses. There were positive effects for 6th grade boys on cooperative classroom climate and major discipline problems; a positive effect for 5th grade girls on locus of control for failure, and a negative effect for 6th grade girls on affective teaching climate. The results were discussed in terms of problems with treatment implementation.

INTRODUCTION

Jigsaw is a structured peer teaching program in which students teach part of the regular classroom curriculum to a group of peers (Aronson, Blaney, Stephan, Sikes & Snapp, 1978). In Jigsaw, an academic lesson is divided into five or six parts and each student in a group of five or six is given only one part of the lesson. Students must become experts on their particular part of the lesson and teach it to students in their groups so that all students learn the entire lesson. Prior to teaching, students meet in Expert groups with their counterparts from other groups. In these groups students help each other prepare to teach their part to their respective Jigsaw groups. Like putting together a jigsaw puzzle, each student contributes a necessary and unique piece of information to group mastery of an assigned task.

The present study investigated the effectiveness of Jigsaw as a strategy to prevent substance abuse. The immediate goals of Jigsaw were to make student learning more active and student-student interactions more cooperative. Reviews of research on cooperative learning (Johnson, 1980b; Sharan, 1980; Slavin, 1980), indicate that these changes should foster positive attitudes, behaviors and norms regarding self, peers and school. Additionally, Jigsaw has positively affected elementary students' self-esteem, attitudes toward peers, and attitude toward school (Blaney, Stephen, Rosenfield, Aronson & Sikes, 1977). Eventually, the predicted positive changes in attitudes and norms should reduce acceptance and use of psychoactive substances (Johnson, 1980a).

Jigsaw in-service training was provided to 13 4th-6th grade teachers in four elementary schools during the 1979-80 school year. The teachers used Jigsaw in their classrooms under the guidance of an in-service trainer. The teachers delivered the intervention; thus, the adequacy of the treatment depended upon teachers' use of Jigsaw. Consequently, teachers' reactions to the training were monitored as were reactions to the conduct of Jigsaw activities in the classroom.

The impact of Jigsaw on both the teachers and their students was assessed. Three teacher outcome variables were measured: teachers' perceptions of the importance of, and their effectiveness at, achieving the objectives of Jigsaw; satisfaction with teaching; and perceived faculty cohesiveness. We hypothesized that Jigsaw would impact all three of these outcome variables.

Student outcome variables were perceived classroom climate, locus of control, social and academic self-esteem, attitudes toward peers, attitudes and perceived peer attitudes toward school, attendance, reading and mathematics achievement, and teacher ratings of student behavior. Drug-specific outcome variables were perceived positive and negative consequences of tobacco, alcohol, and marijuana use; and involvement in use of each of these substances. Among these variables, we hypothesized that Jigsaw would impact all but the drug-related variables. We did not expect effects on the drug-specific variables due to the low level of drug involvement at these ages and the limited duration of the intervention. The present study will be continued for an additional year.

METHOD

Research Design

The subjects in this study participated in another study the prior year. In Fall 1978, eight 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. 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. One school from each pair was randomly assigned to the experimental condition and the other to the control condition. Eighteen of the 23 teachers in grades 4-6 in the experimental schools completed in-service training in Effective Classroom Management (see Schaps, Moskowitz, Condon and Malvin, Note 1, for the procedures and results of this study).

The present study began in Spring 1979. Two of the original experimental schools and two of the original control schools were randomly assigned to the experimental condition, and the other four schools, to the control condition. In Fall 1979, Jigsaw in-service training was offered to the 25 grade 4-6 teachers in the experimental schools. Only 13 teachers completed the training.[1]

[1]Two additional teachers enrolled in the training but dropped out after several weeks. They did not conduct classroom activities related to the training.

Because half of the teachers in the experimental schools did not complete Jigsaw training, about half of the experimental students did not participate in Jigsaw. Comparing all experimentals with all controls provides an insensitive test of Jigsaw's effectiveness. Thus, we have opted to ignore the experimental design and to compare teachers and students who participated in Jigsaw (participants) with all remaining teachers and students (nonparticipants). The research design employs a nonequivalent control group (nonparticipants) with a pretest and a posttest (Cook and Campbell, 1979).

Subjects

The subjects were students in grades 4-6 and their teachers during the 1979-1980 school year. At the beginning of the study, 295 students were enrolled in the 12 participant classes and 685 students were enrolled in the nonparticipant classes. However, 34 (12%) student participants and 125 (18%) student nonparticipants were excluded due to attrition or lack of parental permission for testing.

The participant group consisted of 36, 39, and 38 males, and 56, 42, and 50 females in grades 4-6, respectively. The nonparticipant group consisted of 92, 102, and 85 males, and 107, 93, and 81 females in grades 4-6 respectively. The ethnic composition was 93% white with Mexican-American (3%) comprising the largest minority group.

Of the 13 participant teachers, four were male. Two of the participants taught half-time in the same class. Of the 30 nonparticipant teachers, 12 were male. Four of the nonparticipants taught half-time.

In-Service Training Program

Two-hour in-service training sessions were held once a week for nine weeks. There was an additional two-hour review session six weeks after training ended. During and after the training, the trainer assisted teachers in their classrooms. Teachers who attended the training sessions and tried to implement Jigsaw in their classrooms were paid a $200 stipend and were offered graduate-level credits through a local university.

In the first training session, the trainer provided an overview of the Jigsaw program. Teachers participated in a team-building exercise designed to enhance cooperation, communication, and individual responsibility in groups. They were also provided the materials for implementing this and another exercise in their classrooms. Beginning with the second session, the trainer facilitated weekly discussions of the teachers' experiences introducing the Jigsaw activities in their classrooms.

In the next several sessions teachers learned methods for training students to observe group behavior, to interact with each other productively, and to communicate and listen well.

Sessions five and six focused on developing leadership and problem-solving skills of students. Teachers learned techniques for teaching different leadership styles to students. Teachers also learned specific techniques for effectively intervening in group process to help solve group problems.

In sessions seven through nine, teachers were shown how to develop Jigsaw curricula and were given time to prepare curriculum units for use in their classrooms. Session seven also provided teachers with a method for improving students' teaching skills. Sessions eight and nine included teacher practice

in making observations of group process. (See Tuck and Schaps, Note 2, for training curriculum).

Measures

Student self-report data. Student data at pretest were obtained with an instrument developed for this study called the Student Questionnaire. This instrument consists of two parts. Part 1 contains 18 items selected from the Intellectual Achievement Responsibility Questionnaire (Crandall, Katkovsky, & Crandall, 1965).[2] Selection of items was based upon published psychometric properties. This questionnaire measures the belief in one's own control over intellectual and academic performance. Part 2 contains three sets of items: a) the Scholastic subscale from the Intermediate Level of the Self Appraisal Inventory (20 items) (Instructional Objectives Exchange. Note 3), a criterion-referenced measure of academic self-esteem; b) the Authority and Control (12 items) and Interpersonal Relationships With Pupils (12 items) subscales from the Intermediate-Level of the School Sentiment Index (Instructional Objectives Exchange, Note 4), a criterion-referenced measure of attitudes toward school developed for this study by adapting 11 items from eight instruments that measure attitudes toward school.

Student pretest data were also obtained with the Self Observation Scales (Intermediate Level, Form C) developed by Stenner and Katzenmeyer (Note 5). This instrument measures how children perceive their social and academic selves

---

[2]California Education Code requires that active parental permission be secured before requesting information about family life. Because this study employed passive consent procedures, four items referring to "parents" were adapted to read "an adult who knows you."

and their relationship to peers, teachers, and school. Students respond "yes" or "no" to 60 statements.

Drug-related pretest data were collected using the My Opinion Survey. This instrument was developed for this study to assess student opinions of each of the three "gateway" drugs--alcohol, cigarettes, and marijuana. Specifically, students indicated the consequences of using the drug and reported their lifetime use, their current use ("last four weeks"), their attitudes toward use, their perceptions of peers' attitudes toward use ("most kids in my class"), and their perceptions of peer use.

The pretest measures used in the data analysis were derived from empirical scaling.[3] Pretest data were collected from a large sample of students in grades 3-6. Item means, variances, and intercorrelations were found to be similar across grade levels; thus, students from all three grades were grouped together for the scaling. Item intercorrelations from the pretest data were subjected to separate multiple group confirmatory factor analyses.[4] The resultant pretest scales appear in Table 2 with the number of items and the internal consistency reliability estimated by coefficient alpha. The final scales measured affective teaching climate, attitudes toward school, social self-esteem, attitudes toward peers, locus of control for success, locus of control for failure, academic self-esteem, perceived peer attitudes toward school, perceived costs of and involvement in alcohol, cigarette, and marijuana

[3]The details of the scaling have been reported by Moskowitz, Condon, Brewer, Schaps & Malvin, Note 6.

[4]The correlation matrices were computed using pairwise deletion of missing values and communalities were inserted into their diagonal elements.

TABLE 2

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

Scale	N	Items	Reliability
Affective Teaching Climate (Affec Climate)	513	17	.91
Attitudes Toward School (Att School)	513	6	.74
Social Self-Esteem (Social Self)	513	6	.66
Attitudes Toward Peers (Att Peers)	513	8	.80
Locus of Control: Success (Control Suc)	513	7	.56
Locus of Control: Failure (Control Fail)	513	7	.62
Academic Self-Esteem (Acad Self)	513	11	.75
Perceived Peer Attitudes Toward School (Peer Att Sch)	513	8	.72
Perceived Costs of Alcohol Use (Alc Costs)	386	5	.70
Perceived Costs of Cigarette Use (Cig Costs)	386	5	.63
Perceived Costs of Marijuana Use (Pot Costs)	386	5	.67
Involvement in Alcohol Use (Alc Involve)	386	5	.77
Involvement in Cigarette Use (Cig Involve)	386	5	.75
Involvement in Marijuana Use (Pot Involve)	386	5	.84

use. The involvement scales were comprised of items measuring own use and attitudes, and perceived use and attitudes.[5]

Posttest data were obtained with revised versions of the Student Questionnaire and the My Opinion Survey. The revised instruments included the items in the final pretest measures. The focus of control for success scale contained five additional items and the failure scale contained three additional items. Measures of cooperative (Coop) and competitive (Comp) classroom climate were included in the posttest. These measures, which were developed for this study, assess the students' perceptions of the prevalence of cooperation and competition in their classrooms. The alpha coefficients were .73 and .77, respectively.

Student archival data. In addition to the student self-report data, we used achievement data collected by the school district. These data included the total reading (Read) and total mathematics (Math) stanine scores from the Stanford Achievement Test, Primary Level III (grade 3) and Intermediate Levels I (grade 4) and II (grades 5 and 6) (Madden, Gardner, Rudman, Karlsen, Merwin, 1973), administered in May 1979, and in May 1980.

Student attendance was measured in two ways. 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 occurred when a student was absent from school and did not provide the school with a parental excuse stating that the student was sick. In addition, the average monthly

---

[5]For each substance, the five component items exhibited good internal and external consistency, providing support for a single scale. The positive relationships betweeen the perceived peer measures and the own use and attitude measures indicate that students' own use and attitudes parallel their peers', or that they project their own use and attitudes onto their peers.

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. Both measures were obtained for Spring 1979 and Spring 1980.

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

Teacher data on student behavior. Teacher ratings of student misbehavior were obtained at pretest and posttest with the Student Behavior Report.[6] Using the class roster, each teacher indicated how frequently each child had been a minor (Minor) and major (Major) discipline problem during the previous four months (January through April). A five-point scale was used that ranged from "never" to "about once a day or more."

To determine whether experimental and control teachers used the "minor" and "major" categories in the same manner, teachers rated 17 hypothetical student behaviors as either a) not a discipline problem, b) a minor problem, or c) a major problem. The pretest data were invalid because most teachers did not follow the instructions; hence, instructions for the posttest were revised. Analysis of the posttest data indicated that experimental and control teachers employed similar definitions of minor and major discipline problems.

Teacher outcome data. The teacher's pretest and posttest self-report data were obtained with the Teacher Questionnaire that was developed for this study. This instrument measured a) satisfaction with teaching, adapted from the Purdue Teacher Morale Inventory (Rempel and Bentley, 1964); b) faculty cohesiveness, adapted from the Teacher Cooperation Subscale of the Teacher

[6]The teachers who provided the pretest data were the students' teachers in the prior year.

Attitude and Classroom Climate Questionnaire (Kaufman, Semmel & Agard, Note 7) and from the Intimacy Subscale of the Organization Climate Description Questionnaire (Halpin and Croft, 1963); and c) the importance and effectiveness of achieving teaching objectives related to Jigsaw, a measure developed for this study. Table 3 shows the number of items included in these scales and the pretest internal consistency reliabilities estimated by Coefficient Alpha for grade 3-6 teachers.[7]

Data Collection Procedures

Student survey. Students in grade 4 were given the pretest in October 1979 and the posttest in May 1980. The revised Student Questionnaire was administered at both times. My Opinion Survey was read to the students during a separate session at the pretest.

Students in grades 5 and 6 were given the pretest in May 1979 and the posttest in May 1980. The Student Questionnaire and My Opinion Survey were administered at both times during two sessions in the students' classrooms. The Self Observation Scales were administered at the pretest during the first session. The students read the items to themselves assisted as necessary by the survey administrator. One make up session was held for all students who were absent from the original sessions.

Survey administrators were substitute teachers in the district who were trained by project staff. In a prepared statement, survey administrators assured students of complete confidentiality. The confidentiality of the My Opinion Survey administration was enhanced by prelabeling questionnaires with

---

[7]To increase the sample size, additional teachers were included in the scaling analysis.

TABLE 3

TEACHER SCALES, NUMBER OF ITEMS, AND PRETEST INTERNAL CONSISTENCY RELIABILITIES (COEFFICIENT ALPHA) (N=60)

Scale	Number of Items	Pretest Reliability
Jigsaw Objectives	6	.74
Teacher Satisfaction	8	.80
Faculty Cohesiveness	7	.90

student names on the cover sheet and students' school district identification numbers on page one. After receiving the questionnaire, students were instructed to tear off the cover page which displayed their names. Students were told of the need for identification numbers as a way of tracking students over time.

Teacher survey. Teachers completed the pretest Teacher Questionnaire in September 1978. Teachers completed the posttest questionnaires in their classrooms while their students completed the posttest. To ensure the privacy of their responses, teachers were provided with questionnaires containing unique identifiers.

Data Analysis

The research design employed a nonequivalent control group. With this design, differences obtained between the participant and nonparticipant groups could be unrelated to the treatment. Participating teachers (and their students) were likely to differ from nonparticipants simply because they constituted a special group--teachers who chose to undertake and complete the training. The pretest data were subjected to analysis of variance to explore initial nonequivalence. The posttest data were subjected to analysis of covariance using the corresponding pretest as a covariate. This common technique approximately controls for pre-existing differences. Ideally, each posttest measure should be adjusted for all related pre-existing differences and the pretest measures should be free of measurement error (Reichardt, 1979).

A class-level analysis usually is appropriate with this design. Unfortunately, differences between grade levels were found on many of the variables, and there were few classes in each grade. Therefore, when grade level was

employed as a factor in the design, cell sizes were very small. Thus, statistical analysis performed at the class level had low power. Analysis of student-level data was likely to produce spurious results due to the interdependence of individual students within each class and due to excessive statistical power. We chose to analyze the data at the student level aware that the analysis may provide biased estimates of treatment effects.

In the present study, we set the Type I error rate for each analysis at .05. This is a liberal level given the unit of analysis problem, and the fact that separate analyses were performed for each grade-sex student group due to the existence of heterogeneity of variance and pre-post covariance in many of the measures. Since many analyses were conducted, isolated effects must be interpreted cautiously because they may be due to experiment-wise Type I error. Interpretation of results is based on patterns of findings rather than single findings.

Rules were set for handling missing data in computing 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 cell of the experimental design. This procedure utilized most of the item data and provided unbiased cell means. When more than 40% of the items comprising a scale were missing, the student received a missing value for that scale, and the case was dropped from any analysis involving that scale.

Process Evaluation

Teacher Feedback on the Individual Training Sessions

At the end of each training session (except the final session), teachers were asked to complete a one-page "feedback form" anonymously. Teachers rated each session in terms of organization, interest, usefulness, and enjoyableness on five-point scales (where five was positive). Open ended comments were also encouraged.

All sessions were rated positively. Of the 32 separate ratings (four per session), 26 averaged above 4.50 and only one average rating was less than 4.0. Thus, the teachers regarded the training highly.

Questionnaire Surveys of Participating Teachers

At the completion of training (January 1980), and again near the end of the school year (May 1980), participating teachers completed questionnaires that elicited detailed information about implementation of the in-service skills. Teachers reported how often they used each in-service skill, rated the usefulness of each skill, and rated their mastery of each skill.

Table 4 shows the year-end data on frequency of implementation, usefulness and mastery of the skills. More than half the teachers were using most skills at least several times per week, and two skills every day. Three skills were used by the teachers on less than a weekly basis: conducting group process discussions, coordinating Jigsaw resources, and developing Jigsaw curricula. This implementation pattern indicates that at least half of the participants continued to use the skills for several months after they had completed training.

TABLE 4

TEACHERS' REPORTS ON FREQUENCY, UTILITY, AND QUALITY OF SKILL IMPLEMENTATION (N = 11)

In-Service Skill	Percentage of Teachers Using Skill at Least: Every Day	Several Times/Week	Mean Ratings of Value of Skill (5=High Utility; 1=Low Utility)	Mean Ratings of Mastery of Skills (4=Excellent; 1=Poor)
Observing group process - (communication patterns, decision making procedures, on task/off task behavior, emotional issues)	15	69	4.83	2.92
Giving feedback using the rules for observer feedback	15	69	4.83	2.75
Using attending skills (physically attending, observing, listening)	54	92	4.92	3.08
Responding to feeling	54	92	4.75	3.00
Responding to meaning	39	85	3.67	2.83
Modeling facilitator skills	23	69	4.08	2.58
Intervening to solve group problems using appropriate technique (content intervention, interpersonal intervention, intrapersonal intervention, group intervention)	42	58	4.36	2.60
Conducting group process discussion	15	46	4.23	3.00
Coordinating Jigsaw resources	8	39	3.92	3.00
Developing Jigsaw curricula	8	31	4.00	2.92

Teachers rated all of the skills as extremely useful. The teachers' average ratings of their own mastery of the skills clustered between "fair" and "good." Usefulness and mastery ratings at mid-year were similar to those at year-end.

The mid-year questionnaire 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 (M = 4.54), organization (M = 4.82), usefulness (M = 4.27) and enjoyableness (M = 4.54). The trainer was also rated highly along a number of dimensions.

In sum, results from two questionnaire surveys indicate that many teachers used most of the skills on a regular basis, found the skills useful, considered themselves adept at using the skills, and evaluated the training and the trainer positively.

Frequency of Classroom Implementation

Beginning with the onset of training, and continuing through the school year, participating teachers completed weekly logs indicating the number of classroom hours spent on Jigsaw activities. Data from 12 teachers[8] over 24 weeks indicated wide variation in the implementation of Jigsaw activities; the average number of hours per week each teacher allotted to Jigsaw activities was 1.83 and ranged from .24 to 3.77. Four classrooms were involved in Jigsaw for more than two hours each week, six classrooms were involved in Jigsaw

[8]Data from one teacher were deleted from this analysis because they were obviously falsified.

for 1 to 1.7 hours each week, and one classroom was marginally involved in Jigsaw (.24 hours per week).[9]

Classroom Observation

Observations of Jigsaw activities were conducted three times during the school year; the first observation occurred between the sixth and last week of training, the second visit occurred four to six weeks after training ended, and the final visit occurred eight to 16 weeks after training ended.[10] A random sample of six teachers was observed on the first two visits. On the last visit, all teachers who completed the training, with one exception, were observed, and each visit was scheduled to coincide with the time peer-teaching (Jigsaw groups) would occur.[11]

The observations were conducted by a trained observer. Two instruments were developed to characterize the "Jigsaw classroom," the Jigsaw Checklist and the Jigsaw Observation Record. The Jigsaw Checklist consisted of 13 statements that described the student groups and student interaction. These statements generally were low inference and descriptive rather than evaluative. All statements that were applicable were checked after an observation was

[9]The total number of classrooms counted is one less than the number of teachers because two teachers are part-time in the same classroom. Data from these two teachers were added together and counted as one classroom.

[10]The posttest was administered during the 12th and 13th week after training ended, thus, several classrooms were observed after the posttest.

[11]One teacher who participated in training was not observed because she refused to schedule an appointment with the observer.

completed. Characterizing the groups as Jigsaw groups required that a minimum of five specific statements be checked; also, a specific set of five were used to characterize the groups as Expert groups.

The teachers were expected to be regularly using Jigsaw in their classrooms by the first observation. The data revealed that in only one classroom was Jigsaw implemented as expected. In three of the classrooms, students were engaged in pre-Jigsaw activities or "teambuilding." In the other two classrooms, students were involved in group work that the teachers described as either Jigsaw or Expert, but that did not meet our criteria.

At the second observation, three classrooms were engaged in Jigsaw or Expert group work. In three other classrooms, although the teachers told the observer that their students were in Jigsaw groups, the observer found no evidence for this. These teachers had created their own versions of Jigsaw. The critical difference in their versions of Jigsaw was that the students were working independently rather than cooperatively on an assignment, although they were seated in groups.

Because Jigsaw is designed to be a peer-teaching technique, we examined the extent to which the teaching component was implemented. For the third series of observations, all teachers were observed during the time their students were in Jigsaw groups. Of the 12 classrooms observed, five demonstrated the expected Jigsaw activities prior to the posttest. Two additional classrooms implemented the peer-teaching aspect of Jigsaw before the end of the year, but after the posttest data were collected. These classes were involved in Expert groups before the posttest. The remaining five classrooms

were involved in various modifications of Jigsaw. These modifications had at least one feature in common; students were not teaching each other in small groups but were working independently on a given assignment.

In two of the five true Jigsaw classrooms observed before the posttest, and in one true Jigsaw classroom observed afterwards, student groups were off-task for a substantial amount of time. A serious task orientation is necessary for adequate Jigsaw implementation; therefore, these classrooms did not fully meet our expectations. Three true Jigsaw classrooms observed before the posttest, and one observed afterwards, were exemplary in terms of implementation. These classrooms not only met the minimum Jigsaw criteria but were notable for the focussed attention of the students on the tasks at hand.

The Jigsaw Observation Record was designed to provide more specific detail regarding student behavior during group work. This instrument employed time samples primarily of student behavior. Each group was observed for 30 seconds, and all the codeable behaviors that occurred during that interval were recorded. A behavior was recorded only once, even though it may have occurred several times during the observation period. Each group was observed approximately five times during one visit. Since all the participants were observed during the third round of observations, only the coded behavior from this series of visits is reported below.

The amount and type of student participation in the groups was approximately the same for all classrooms. A student leader could clearly be distinguished in all the classrooms across most of the intervals of observation. There was very little helping behavior overall, with the exception of one classroom. Silent group work was observed frequently. The behavior "shares ideas" was

coded whenever a student shared information with the group and, thus, included all peer-teaching behaviors. The frequency with which this behavior occurred discriminated two groups of classrooms. In one group of classrooms, this behavior occurred frequently. In these classrooms, Jigsaw was accurately implemented as indicated by the Checklist. In the other classrooms, relatively little student teaching was observed, and modified versions of Jigsaw were used in them.

Off-task and disruptive behaviors were observed in both Jigsaw and modified Jigsaw classrooms. However, there was virtually no off-task behavior in five classrooms, four of which were Jigsaw classrooms. Finally, "teacher helps" was coded whenever the teacher became involved with a group upon request of a group member. This occurred almost exclusively in classrooms employing modifications of Jigsaw. Students were more inclined to seek out the teachers' help in variations of Jigsaw than they were in true Jigsaw.

RESULTS

Initial Equivalence--Teachers

Teacher self-report data collected prior to the training were analyzed to determine whether participants and nonparticipants were initially equivalent. The results indicated that the groups did not differ in terms of teaching experience, use of various academic and affective techniques in the classroom, Teacher Satisfaction or Faculty Cohesiveness. However, compared to non-participants, participants scored higher on Jigsaw Objectives, $F(1,23) = 4.80$, $p < .05$. These results should be interpreted cautiously because pretest data were missing from 42% of the teachers.

The groups differed on prior in-service training. Sixty-two percent of the participants completed ECM in-service training in the prior year as compared to only 23% of the nonparticipants.

Analysis of Teacher Outcomes

The three teacher posttest measures were subjected to one-way (participant vs. nonparticipant) analyses of variance. No significant differences were obtained.

Descriptive Statistics for Student Measures

The means and standard deviations for all student pretest and posttest measures are summarized in Tables 5-7 by grade, sex, and treatment condition. Due to extreme non-normality and heterogeneity of variance, log (x+1) transformations were performed on the following variables: Attend, Unex Abs,

TABLE 5

MEANS AND STANDARD DEVIATIONS FOR STUDENT DATA BY SEX BY TREATMENT CONDITION FOR GRADE 4

		Male Students				Female Students			
		Participant		Nonparticipant		Participant		Nonparticipant	
Measure	Test	M	SD	M	SD	M	SD	M	SD
Coop	Post	1.71	.26	1.67	.26	1.83	.21	1.74	.28
Comp	Post	1.56	.29	1.50	.31	1.52	.31	1.47	.28
Affec Climate	Pre	1.85	.18	1.86	.20	1.93	.12	1.87	.20
	Post	1.80	.20	1.79	.28	1.90	.14	1.81	.27
Att School	Pre	1.81	.28	1.80	.22	1.79	.24	1.78	.23
	Post	1.69	.31	1.66	.29	1.80	.21	1.71	.29
Social Self	Pre	1.81	.28	1.80	.22	1.79	.24	1.78	.23
	Post	1.83	.22	1.79	.21	1.80	.26	1.77	.24
Att Peers	Pre	1.68	.25	1.73	.26	1.69	.25	1.71	.27
	Post	1.75	.31	1.71	.28	1.69	.28	1.70	.30
Control Suc	Pre	1.80	.15	1.81	.15	1.84	.11	1.84	.14
	Post	1.76	.17	1.77	.19	1.86	.14	1.83	.16
Control Fail	Pre	1.55	.25	1.61	.24	1.64	.22	1.64	.24
	Post	1.65	.25	1.66	.22	1.67	.23	1.67	.24
Acad Self	Pre	1.72	.23	1.74	.21	1.76	.20	1.75	.21
	Post	1.72	.23	1.72	.23	1.77	.22	1.73	.24
Peer Att Sch	Pre	1.74	.20	1.75	.23	.82	.16	.81	.21
	Post	1.70	.23	1.68	.25	1.75	.18	1.70	.27
Total Abs	Pre	-	-	-	-	-	-	-	-
	Post	.27	.19	.23	.19	.33	.24	.27	.20
Unex Abs	Pre	-	-	-	-	-	-	-	-
	Post	.13	.24	.18	.27	.15	.20	.15	.24
Read	Pre	-	-	-	-	-	-	-	-
	Post	6.29	1.68	5.89	1.76	5.91	1.48	6.27	1.30
Math	Pre	-	-	-	-	-	-	-	-
	Post	5.88	1.72	5.98	1.67	5.44	1.30	5.97	1.38

Table 5

Measure	Test	Male Students				Female Students			
		Participant		Nonparticipant		Participant		Nonparticipant	
		M	SD	M	SD	M	SD	M	SD
Minor	Pre	-	-	-	-	-	-	-	-
	Post	2.47	1.56	1.97	1.33	1.64	1.09	1.22	.63
Major	Pre	-	-	-	-	-	-	-	-
	Post	.33	.68	.33	.10	.31	.03	.31	.05
Alc Costs	Pre	-	-	-	-	-	-	-	-
	Post	.56	.48	.54	.50	.45	.40	.59	.49
Cig Costs	Pre	-	-	-	-	-	-	-	-
	Post	.48	.47	.48	.45	.38	.36	.47	.41
Pot Costs	Pre	-	-	-	-	-	-	-	-
	Post	.46	.37	.42	.46	.39	.36	.39	.39
Alc Involve	Pre	-	-	-	-	-	-	-	-
	Post	1.54	.69	1.37	.63	1.30	.61	1.40	.58
Cig Involve	Pre/	-	-	-	-	-	-	-	-
	Post	.27	.07	.27	.08	.26	.06	.26	.07
Pot Involve	Pre	-	-	-	-	-	-	-	-
	Post	.25	.06	.24	.07	.24	.06	.25	.08

TABLE 6

MEANS AND STANDARD DEVIATIONS FOR STUDENT DATA BY SEX BY TREATMENT CONDITION FOR GRADE 5

		Male Students				Female Students			
		Participant		Nonparticipant		Participant		Nonparticipant	
Measure	Test	M	SD	M	SD	M	SD	M	SD
Coop	Post	1.69	.30	1.71	.26	1.69	.30	1.78	.26
Comp	Post	1.50	.32	1.48	.31	1.36	.25	1.42	.28
Affec Climate	Pre	1.81	.09	1.70	.21	1.82	.08	1.75	.18
	Post	1.85	.21	1.80	.26	1.84	.26	1.90	.16
Att School	Pre	1.73	.27	1.64	.32	1.76	.28	1.74	.27
	Post	1.72	.29	1.62	.29	1.72	.25	1.73	.24
Social Self	Pre	1.78	.23	1.78	.26	1.75	.24	1.79	.25
	Post	1.85	.23	1.84	.21	1.85	.19	1.85	.23
Att Peers	Pre	1.73	.29	1.71	.30	1.77	.23	1.72	.28
	Post	1.83	.25	1.74	.28	1.70	.29	1.69	.29
Control Suc	Pre	1.85	.19	1.80	.21	1.89	.13	.88	.15
	Post	.83	.19	1.81	.19	1.85	.11	1.83	.18
Control Fail	Pre	1.70	.20	1.67	.27	1.71	.22	1.72	.25
	Post	1.71	.22	1.71	.23	1.78	.19	1.65	.26
Acad Self	Pre	1.77	.20	1.74	.26	1.81	.16	1.76	.25
	Post	1.75	.25	1.76	.21	1.74	.22	1.76	.25
Peer Att Sch	Pre	1.79	.21	1.70	.28	1.75	.20	1.78	.21
	Post	1.76	.20	1.70	.24	1.77	.24	1.74	.26
Total Abs	Pre	.21	.17	.18	.16	.25	.20	.22	.19
	Post	.20	.18	.25	.16	.29	.21	.28	.19
Unex Abs	Pre	.17	.30	.24	.28	.15	.26	.27	.30
	Post	.15	.22	.14	.25	.12	.23	.19	.29
Read	Pre	5.44	1.60	5.83	1.53	5.71	1.45	6.24	1.43
	Post	5.85	1.60	6.21	1.45	6.12	1.29	6.29	1.43
Math	Pre	5.69	1.59	5.88	1.63	5.31	1.44	6.00	1.55
	Post	6.10	1.57	5.98	1.59	5.87	1.38	5.83	1.31

Table 6

Measure	Test	Male Students				Female Students			
		Participant		Nonparticipant		Participant		Nonparticipant	
		M	SD	M	SD	M	SD	M	SD
Minor	Pre	1.82	.79	1.97	1.12	2.08	1.27	1.27	.63
	Post	2.12	1.24	2.09	1.30	1.40	.86	1.41	.80
Major	Pre	.37	.14	.36	.12	.33	.08	.32	.09
	Post	.32	.06	.37	.12	.30	.03	.32	.09
Alc Costs	Pre	.53	.74	.49	.50	.49	.37	.46	.43
	Post	.45	.54	.57	.56	.48	.43	.40	.45
Cig Costs	Pre	.32	.32	.35	.35	.46	.43	.41	.36
	Post	.35	.47	.45	.47	.40	.44	.41	.41
Pot Costs	Pre	.29	.30	.34	.40	.41	.37	.36	.35
	Post	.25	.36	.42	.50	.35	.42	.31	.37
Alc Involve	Pre	1.40	.66	1.40	.64	1.20	.44	1.18	.50
	Post	1.50	.75	1.64	.78	1.43	.56	1.38	.59
Cig Involve	Pre	.25	.05	.25	.06	.25	.07	.24	.06
	Post	.26	.07	.30	.10	.27	.08	.30	.86
Pot Involve	Pre	.22	.04	.23	.06	.22	.04	.23	.05
	Post	.22	.03	.26	.10	.24	.07	.25	.07

TABLE 7

MEANS AND STANDARD DEVIATIONS FOR STUDENT DATA BY SEX BY TREATMENT CONDITION FOR GRADE 6

		Male Students				Female Students			
		Participant		Nonparticipant		Participant		Nonparticipant	
Measure	Test	M	SD	M	SD	M	SD	M	SD
Coop	Post	1.83	.21	1.65	.29	1.84	.24	1.75	.26
Comp	Post	1.38	.29	1.41	.29	1.40	.31	1.41	.28
Affec Climate	Pre	1.74	.19	1.71	.19	1.80	.11	1.78	.13
	Post	1.74	.31	1.79	.28	1.69	.34	1.87	.18
Att School	Pre	1.71	.29	1.59	.30	1.75	.27	1.74	.27
	Post	1.59	.27	1.59	.30	1.64	.26	1.68	.28
Social Self	Pre	1.88	.16	1.85	.19	1.80	.27	1.83	.22
	Post	1.86	.19	1.87	.17	1.81	.23	1.86	.20
Att Peers	Pre	1.74	.27	1.75	.27	1.72	.30	1.72	.28
	Post	1.77	.29	1.78	.27	1.71	.33	1.76	.26
Control Suc	Pre	1.86	.19	1.82	.18	1.81	.22	1.87	.17
	Post	1.83	.16	1.82	.18	1.83	.17	1.88	.12
Control Fail	Pre	1.69	.26	1.70	.24	1.76	.25	1.81	.23
	Post	1.75	.21	1.73	.22	1.76	.23	1.82	.18
Acad Self	Pre	1.79	.20	1.74	.22	1.68	.30	1.77	.22
	Post	1.75	.20	1.74	.23	1.63	.30	1.76	.22
Peer Att Sch	Pre	1.77	.24	1.68	.24	1.80	.22	1.80	.24
	Post	1.68	.24	1.63	.27	1.67	.33	1.61	.29
Total Abs	Pre	.21	.20	.23	.18	.20	.19	.24	.19
	Post	.21	.16	.25	.17	.26	.23	.26	.20
Unex Abs	Pre	.27	.31	.23	.30	.15	.26	.27	.30
	Post	.18	.27	.14	.26	.16	.25	.14	.25
Read	Pre	5.91	1.40	5.96	1.57	5.86	1.35	6.05	1.52
	Post	5.76	1.51	6.01	1.64	6.14	1.54	6.38	1.72
Math	Pre	5.63	1.41	5.83	1.64	5.31	1.44	6.00	1.55
	Post	5.79	1.61	5.75	1.69	5.69	1.83	6.00	1.56

Table 7

Measure	Test	Male Students				Female Students			
		Participant		Nonparticipant		Participant		Nonparticipant	
		M	SD	M	SD	M	SD	M	SD
Minor	Pre	2.54	1.50	2.42	1.33	2.08	1.27	1.27	.63
	Post	2.35	1.36	2.46	1.42	1.82	.80	1.69	1.11
Major	Pre	.35	.12	.33	.09	.34	.11	.30	.02
	Post	.35	.10	.42	.16	.31	.04	.32	.09
Alc Costs	Pre	.57	.50	.60	.50	.31	.32	.48	.40
	Post	.76	.52	.63	.56	.38	.43	.48	.42
Cig Costs	Pre	.33	.32	.40	.43	.39	.35	.46	.43
	Post	.48	.36	.59	.46	.41	.46	.54	.46
Pot Costs	Pre	.49	.45	.36	.50	.28	.29	.31	.32
	Post	.48	.41	.55	.56	.29	.42	.35	.40
Alc Involve	Pre	1.42	.73	1.51	.62	1.20	.44	1.18	.50
	Post	1.84	.70	1.87	.69	1.43	.55	1.40	.46
Cig Involve	Pre	.25	.09	.28	.09	.27	.08	.27	.07
	Post	.30	.08	.33	.10	.31	.10	.33	.09
Pot Involve	Pre	.23	.06	.25	.08	.22	.04	.23	.05
	Post	.26	.07	.29	.10	.26	.08	.27	.07

Major, Cig, and Pot. For attendance, discipline, and drug-related variables, a low score is desirable; whereas, for all other variables a high score is desirable.

Statistical analysis of measures with floor or ceiling effects generally suffers from decreased power (due to decreased reliability) and from bias. For example, students who are at the ceiling on a pretest measure where a high score is desirable cannot show improvement on the identical measure at posttest. This could bias an analysis against finding a positive treatment effect.

Ceiling effects were found on some of the affective measures. The percentage of students with a maximum score at the pretest ranged from 8% on Comp to 42% on Social Self with a median value of 26%. The corresponding percentages at the posttest ranged from 7% to 46% with a median value of 21%.

On the drug-related measures, evidence was found for floor effects. The percentage of students with a minimum score at the pretest ranged from 16% on Alc Involve to 69% on Pot Involve with a median value of 35%. The corresponding percentages at the posttest ranged from 7% to 49% with a median value of 27%.

In sum, the affective measures did not suffer from substantial ceiling effects. Although floor effects were substantial on several of the drug-related measures, they were expected given the age of the students. Furthermore, they do not constitute a problem for the present study, because no direct effects on these measures were hypothesized.

Initial Equivalence--Students

Initial equivalence between conditions affects both the justification for attributing posttest differences to the treatment (internal validity)

and the justification for generalizing treatment effects to other student groups (external validity). As a test for initial equivalence, analysis of variance was performed on each of the 14 (8 for grade 4) student pretest measures with treatment condition (participant vs. nonparticipant) as the factor. Effect sizes, expressed in the metric of the nonparticipant group standard deviation, are presented for measures with significant differences.

For grade 4 males no significant treatment differences were obtained. For grade 4 females treatment differences were obtained for Att School, F (1,61) = 5.09, p <.05, and Affec Climate, F (1,160) = 4.67, p <.05. As compared to nonparticipants, participants were initially more positive toward school (.34 SD) and more positive toward the teaching climate (.31 SD). A treatment difference was obtained for grade 5 males on Affec Climate, F (1,137) = 9.75, p <.01. As compared to nonparticipants, participants were more positive toward the teaching climate (.52 SD). For grade 5 females, treatment differences were obtained on Affec Climate, F (1,124) = 4.67, p <.05, Read, F (1,129) = 3.93, p < .05, and Math F (1,129) = 9.28, p < .05. As compared to nonparticipants, participants were more positive toward the teaching climate (.37 SD) and scored lower on tests of reading (.37 SD) and mathematics (.39 SD) achievement. A treatment difference was obtained for grade 6 males on Att School, F (1,117) = 4.57, p <.05. As compared to nonparticipants, participants were more positive toward school (.41 SD). Finally, for grade 6 females, treatment differences were obtained on Unex Abs, F (1,128) = 5.46, p <.05, Math, F (1.18) = 5.71, p <.05, Minor, F (1,119) = 21.67, p <.001, Major, F (1,120) = 6.92, p <.01, and Alc Costs, F (1,121) = 5.91, p <.05.

As compared to nonparticipants, participants had fewer unexcused absences (.40 SD) and lower mathematics achievement scores (.45 SD). Participants also presented more minor (1.29 SD) and major (1.64) discipline problems, and attributed fewer negative consequences to alcohol use (.41 SD).

In sum, with the exception of grade 4 males, participants initially differed from nonparticipants; the differences occurred on more variables for females than for males.

Analysis of Student Outcomes

A one-way analysis of covariance was conducted on each posttest measure with treatment condition (participant vs. nonparticipant) as the factor and the corresponding pretest as the covariate. Because Coop and Comp were not measured at pretest, analyses of variance were conducted on the posttest data. For grade 4, analyses of variance were also conducted on the following posttest measures due to the lack of pretests: Attend, Unex Abs, Minor, Major, Read, Math, and the drug-related variables. Treatment effect sizes are presented in the metric of the posttest nonparticipant group standard deviation.

Table 8 summarizes the results of the posttest analyses. For grade 4 males, a significant positive treatment effect was obtained on Att Peers (.29 SD). For grade 4 females, negative treatment effects were obtained on Minor (.67 SD) and Math (.39 SD). For grade 5 males, positive treatment effects were obtained on Major (.49 SD), Cig (.40 SD), and Pot (.41 SD). For grade 5 females, positive treatment effects were obtained on Control Fail (.50 SD) and Cig (.35 SD), and a negative effect on Math (.28 SD). For grade 6 males, positive treatment effects were obtained on Coop (.57 SD) and Major (.44 SD). For grade 6 females positive treatment effects were

TABLE 8

SUMMARY OF ANALYSES OF COVARIANCE ON STUDENT DATA

Measure	Grade 4 Males	Grade 4 Females	Grade 5 Males	Grade 5 Females	Grade 6 Males	Grade 6 Females
Coop	1.60	3.34	<1	3.15	10.30**	4.74*
Comp	<1	2.97	<1	1.72	<1	<1
Affec Climate	<1	2.55	<1	<1	1.52	13.80**
Att School	2.63	1.65	1.66	<1	<1	<1
Social Self	<1	<1	<1	<1	<1	<1
Att Peers	4.47*	<1	2.15	1.13	<1	<1
Control Suc	<1	2.37	<1	<1	<1	1.06
Control Fail	<1	<1	<1	10.09**	<1	1.02
Acad Self	<1	1.09	<1	2.02	<1	3.63
Peer Att Sch	<1	1.99	<1	1.52	<1	2.35
Total Abs	1.28	3.09	3.70	<1	1.73	<1
Unex Abs	<1	<1	<1	1.62	<1	1.44
Read	1.34	2.47	<1	3.06	<1	<1
Math	<1	5.34*	1.88	6.13*	3.38	3.99*
Minor	3.37	9.67**	<1	<1	<1	1.85
Major	<1	<1	8.32**	2.92	8.25**	1.61
Alc Costs	<1	3.24	2.33	<1	2.06	<1
Cig Costs	<1	1.98	<1	<1	1.31	2.61
Pot Costs	<1	<1	2.72	<1	<1	<1
Alc Involve	1.68	1.06	1.51	<1	<1	<1
Cig Involve	<1	<1	6.16*	4.61*	<1	<1
Pot Involve	<1	<1	7.66**	<1	2.09	<1

NOTE: The values tabled are the F-ratios for the treatment condition effect. Error df are 122-126 for grade 4 males, 155-161 for grade 4 females, 110-138 for grade 5 males, 109-132 for grade 5 females, 108-120 for grade 6 males, and 117-128 for grade 6 females.

* = $p < .05$

obtained on Coop (.39 SD), and Math (.27 SD), and a negative effect on Affec. Climate (.87 SD).

Analysis of Student Outcomes for "Exemplary" Jigsaw

The student outcome analysis did not yield a pattern of treatment effects. The results could possibly be due to the variation in frequency and quality of Jigsaw implementation (Cook and Campbell, 1979). A secondary analysis was performed in order to determine whether the "exemplary" Jigsaw classes obtained a pattern of student outcomes. Three exemplary Jigsaw classes were identified using the Jigsaw Checklist data--one 5th, one 6th, and one 5th-6th grade. combination. The classes spent an average of 2.4 hours per week engaged in Jigsaw.

A one-way analysis of covariance was conducted on each posttest measure with condition (exemplary Jigsaw participant vs. nonparticipant) as the factor and the corresponding pretest as the covariate. Analyses of variance were conducted on the Coop and Comp posttests. Separate analyses were performed for each grade-sex group. The significant results are summarized below. For grade 5 males, exemplary Jigsaw participants had better attendance (Attend) F (1,109) = 4.24, p <.05. For grade 5 females, exemplary participants attributed failure more internally (Control Fail), F (1,99) = 5.31, p <.05, perceived more positive peer attitudes toward school (Peer Att School), F (1,101) = 9.71, p < .01 and perceived their classrooms as less competitive (Comp), F (1,103) = 8.34, p < .01. For grade 6 males, exemplary participants perceived their classrooms as more cooperative (Coop), F (1,102) = 7.28, p < .01, perceived more positive peer attitudes toward school, F (1,99) = 5.22, p <.05, and were reported by their teachers to engage in fewer major discipline

problems (Major), F (1,96) = 7.08, p <.01. Finally, for grade 6 females, exemplary participants perceived more positive peer attitudes toward school, F (1,94) = 6.88, p <.01, and were less involved with cigarettes (Cig Involve), F (1,97) = 7.08, p <.01, and marijuana use (Pot Involve), F (1,98) = 8.56, p <.01; however, they also perceived a more negative teaching climate (Affec Climate), F (1,96) = 5.11, p <.05, and had more unexcused absences (Unex Abs), F (1,98) = 6.24, p <.05.

The results from these analyses also do not form a pattern. The positive effect on Peer Att School which was found for three of the four subgroups did not replicate in the primary analyses which compared all participants with non-participants. Of the 26 significant effects obtained in both sets of analyses for grades 5 and 6, there were only four effects common to both: a positive effect on Control Fail for grade 5 females, positive effects on Coop and Major for grade 6 males, and a negative effect on Affec Climate for grade 6 females.

Discussion

The present study shares some methodological weaknesses with previous studies of Jigsaw. Most studies have used designs in which units were assigned nonrandomly to condition. Furthermore, all studies employed student level analyses, whereas classrooms were the units treated. This analysis strategy creates a risk of obtaining spurious results because it inflates statistical power and fails to control for bias due to student interdependence. The present outcome results should be interpreted in light of these important limitations.

The hypothesized effects of Jigsaw were not supported by the data. There were no significant differences on the teacher measures of faculty cohesiveness, teaching satisfaction, or the importance and effectiveness of achieving teaching objectives related to Jigsaw. Although there were several effects on students' attitudes and behavior, these generally were not obtained across grade-sex groups or across primary and secondary analyses. Some isolated effects that replicated in both analyses were the positive effects for 6th grade boys on cooparative classroom climate and major discipline problems, the positive effect for 5th grade girls on locus of control for failure and the negative effect for 6th grade girls on affective teaching climate.

Peer teaching programs have been used by educators and prevention practitioners to increase students' satisfaction with school, teachers and peers, and to improve self-esteem. Jigsaw was selected because it is a peer teaching program that involves all students. Jigsaw is different from many peer teaching

programs in that it provides opportunities for each student to be both teacher and pupil in the context of a small, interdependent work group.

The Jigsaw format dictates specific conditions that may have been inadequately met in the present study. In Jigsaw classrooms, children must depend on one another to learn the regular curriculum. Student interdependence is required for learning, and dependence on the teacher or on the smartest students is avoided. In the present study, trainer anecdotes, based on classroom observation, indicated that many teachers seemed reluctant to allow their student groups to work autonomously. The researcher's observations were consistent with this since teacher interference in student groups was frequent in half of the experimental classrooms. It seems likely that student interdependence was undermined by certain intrusive teacher behaviors.

The element of "required" interdependence among students is what makes Jigsaw a distinctive peer teaching program, and yet, it was probably the most poorly implemented aspect of the program. Students are interdependent to the extent that each child has a unique and critical piece of information to contribute to group learning of an assigned task. In about 60% of the participating classrooms, students had limited or no opportunity to contribute to group learning. In a few classrooms, students spent an excessive amount of time in their expert groups which delayed teaching in Jigsaw groups until after the posttest. Jigsaw was modified by the teachers and student teaching was eliminated in several other classrooms. Where student teaching was eliminated, students completed worksheets and simply read or held up the answers for other students in their groups to copy. In these classrooms, students were working individually or in pairs as opposed to working cooperatively on a topic.

Apparently, students were not well prepared to work with one another productively. The observational data suggest that in half the classrooms there was a striking lack of constructive group behavior. According to Aronson, et al. (1978), it is essential that students learn cooperative, constructive group process skills before they actually work with academic material. They recommend that teachers involve students in team-building exercises for a short period each day over one or two weeks so that students will learn mutual respect and interpersonal skills. Both the observational data and the teacher logs provided evidence that team-building activities were conducted much less intensively than recommended.

The literature on Jigsaw provides only a vague notion regarding minimal effective levels of Jigsaw implementation. Beneficial effects of Jigsaw have been reported with two to four hours of Jigsaw per week. The lack of consistent positive effects in the present study does not seem attributable to inadequate frequency of implementation. Considering only the exemplary teachers' classrooms, we found that about 2.4 hours per week, or 11% of student time in the classroom was devoted to Jigsaw. This finding compares favorably, in terms of number of hours per week, with implementation levels reported by Blaney, et al. (1977).

It must be pointed out that there were a number of differences between the present study and previous research on Jigsaw. For example, previous studies were short-term; teachers were asked to implement Jigsaw in their classrooms for only two to eight weeks. In the present study, teachers were asked to implement Jigsaw for approximately six months. On the positive side, the longer duration of this study should have produced more beneficial effects than found by Blaney, et al. since students were exposed to more total hours

of Jigsaw, particularly in our exemplary classrooms. Curiously, even in these three classrooms, we did not find the predicted positive effects. For the other participating teachers, the length of involvement in this study may have seemed costly in terms of time and effort, which may have impacted the quality or frequency of Jigsaw implementation.

This study may simulate more closely than previous studies the impact of Jigsaw under naturalistic conditions. Although careful monitoring of program implementation is atypical, teacher in-service training is often selected or imposed by school district officials and conducted by outsiders. Where teachers do not themselves initiate a request for Jigsaw training, their reception of Jigsaw may not meet the expectations of those providing the training. Many teachers may opt not to participate in the training (in this study, about one-half of the eligible teachers), or they may participate without much commitment. Of those teachers who complete the training, many may modify Jigsaw by eliminating critical components (e.g., peer teaching or task interdependence) thereby undermining it's effectiveness. Few teachers may actually implement Jigsaw appropriately and effectively in their classrooms (in this study, about one-fourth of those who completed the training).

This speculation may be unduly pessimistic, especially since our recruitment, training, and classroom follow-up procedures all could be improved. Nevertheless, many teachers may be reluctant to make the necessary change in role from instructor to facilitator regardless of the nature or amount of in-service training. Future research should be conducted to determine how to overcome teacher resistance and improve classroom implementation. To the extent possible, further studies of Jigsaw should employ experimental designs with classes as the unit of analysis, and include extensive process evaluation.

Reference Notes

1. Schaps, E., Moskowitz, J., Condon, J., & Malvin, J. A Process and Outcome Evaluation of an Affective Teacher Training Prevention Program, Prevention Branch, National Institute on Drug Abuse, 1980.

2. Tuck, P., & Schaps, E. Jigsaw Training Manual, Prevention Branch, National Institute on Drug Abuse, 1981.

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

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

5. Stenner, A., & Katzenmeyer, W. Technical Manual and User's Guide: Self Observation Scales, NTS Research Corp., Durham, N.C., undated.

6. 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.

7. 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.

REFERENCES

Aronson, E., Blaney, N., Stephan, C., Sikes, J., & Snapp, M. *The Jigsaw Classroom*, Beverly Hills, Sage Publications Inc., 1978.

Blaney, N., Stephan, C., Rosenfield, D., Aronson, E., & Sikes, J. Interdependence in the classroom: A field study, *Journal of Educational Psychology*, 1977, *69*, 121-128.

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.

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

Johnson, D. Constructive peer relationships, social development, and cooperative learning experiences: Implications for the prevention of drug abuse. *Journal of Drug Education*, 1980a, *10*, 7-24.

Johnson, D. Group processes: Influences of student-student interaction on school outcomes. In McMillan, J. (Ed.), *The Social Psychology of School Learning*, Academic Press, New York, 1980b, 123-168.

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

Reichardt, C. The Statistical Analysis of Data from Nonequivalent Group Designs. In Cook, T. and Campbell, D. (Eds.), *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Chicago: Rand McNally, 1979.

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

Sharan, S. Cooperative learning in small groups: Recent methods and effects on achievement, attitudes, and ethnic relations. *Review of Educational Research*, 1980, *50*, 241-271.

Slavin, R. Cooperative learning. *Review of Educational Research*, 1980, *50*, 315-342.