Evaluation of Jigsaw, a Cooperative Learning Technique

JOEL M. MOSKOWITZ, JANET H. MALVIN, GARY A. SCHAEFFER, AND ERIC SCHAPS
Prevention Research Center, Berkeley, California

Cooperative learning techniques have been promoted for the development of academic and social competencies. One such technique, Jigsaw, creates cooperation by structuring student interdependence through the learning task, rather than through the grading system. A process and outcome evaluation of Jigsaw was conducted. Eleven teachers of fifth-grade classes received Jigsaw in-service training and conducted Jigsaw in their classes over a school year. Students in 13 other fifth grade classes served as a comparison group. Students received a pretest and a post-test assessing attitudes toward self, peers, and school, and achievement and attendance records were collected. The process evaluation revealed that the quality and frequency of Jigsaw implementation varied greatly. Jigsaw failed to have a positive effect on the outcome variables, even for the five classes where it was implemented proficiently. The results, which are consistent with an earlier study (J. Moskowitz, J. Malvin, G. Schaeffer, & E. Schaps, 1983, American Education Research Journal, 20, 687-696), are discussed in terms of a theoretical shortcoming of this technique. © 1985 Academic Press, Inc.

Over the past decade, researchers have investigated the utility of classroom cooperative learning for school desegregation (Aronson, Blaney, Stephan, Sikes, & Snapp, 1978), for substance abuse prevention (Moskowitz, Malvin, Schaeffer, & Schaps, 1983), and for mainstreaming academically handicapped students (Madden & Slavin, 1983). Various cooperative learning techniques have been developed and evaluated in several dozen field studies. As compared to traditional instructional strategies, cooperative learning techniques improve students' attitudes toward themselves, their peers, and their school, in addition to enhancing academic performance (Johnson, 1980; Johnson, Maruyama, Johnson, Nelson, & Skon, 1981; Sharan, 1980; Slavin, 1980).

Jigsaw is a cooperative learning technique in which students teach part of the regular curriculum to a small group of their peers (Aronson et al., 1978). The classroom teacher divides a lesson into five or six parts and

This study was supported by a grant from the Prevention Branch, National Institute on Drug Abuse (DA02147). The final version of this paper was written with the support of a National Institute on Alcohol Abuse and Alcoholism research center grant (AA06282-01). The opinions stated in this report are those of the authors. Address reprint requests to Joel M. Moskowitz, PhD, Prevention Research Center, 2532 Durant Avenue, Berkeley, CA 94704.
JIGSAW EVALUATION

105

gives one part of the lesson to each student in a Jigsaw group of five or six students. Prior to teaching in Jigsaw groups, students meet in "expert" groups with their counterparts from other Jigsaw groups to help each other prepare to teach their part to their respective Jigsaw groups. In the Jigsaw group, each student teaches a necessary and unique piece of information to help the group master the assigned work. When the unit is completed, the students are tested and they each receive a grade based upon their own test performance. This contrasts with other cooperative learning techniques in which students receive a grade based upon their group performance. With Jigsaw, student cooperation is promoted by structuring student interdependence through the learning task rather than through the grading system.

Prior evaluations of Jigsaw showed mixed results. Lucker, Rosenfield, Sikes, and Aronson (1976) found that, as compared to traditional classrooms, Jigsaw improved academic performance on a social studies lesson for blacks and Mexican-Americans, but not for Anglos. Blaney, Stephan, Rosenfield, Aronson, and Sikes (1977) found that Jigsaw participants had greater self-esteem and liked students in their group better, but did not have more positive attitudes toward their classmates. In this study, Jigsaw also had a positive effect on attitudes toward school for blacks and Anglos, but had a negative effect for Mexican-Americans. Geffner (no date) found that Jigsaw had a positive effect on general self-esteem, but had both positive and negative effects on different measures of academic self-esteem and attitudes toward peers. Jigsaw had a negative effect on attitudes toward school for Anglo females, but not for other subgroups. Finally, our prior evaluation of Jigsaw found no effects on academic achievement, self-esteem, attitudes toward school, attitudes toward peers, or locus of control (Moskowitz et al., 1983).

Prior evaluations of Jigsaw have suffered from methodological limitations which may have contributed to the inconsistent results. All of the studies employed weak research designs or inappropriate statistical analyses. Lucker et al. and Blaney et al. utilized nonequivalent control group designs in which Jigsaw teachers chose their control group counterparts. Furthermore, Blaney et al. discarded two of their five control classes. Geffner (no date) randomly assigned classes to condition, but he only had four Jigsaw classes and two control classes. In our prior study, schools were randomly assigned to condition, but only half of the teachers who were eligible for Jigsaw participated; thus, we utilized a randomized invitation design (Brewer, 1976). With the exception of Lucker et al., the data analyses in previous studies assumed that the students within each class were statistically independent. Although this is a commonly made assumption, it is generally an unwarranted one, and when it is violated, the tests of significance are misleading (Hopkins, 1982).
With the exception of our prior Jigsaw evaluation, previous studies were of short duration (2 to 8 weeks) and did not conduct process evaluations. Thus, prior results could be due to inadequate implementation of this learning technique. Our prior findings do not support this explanation, however, as we did not find any pattern of effects even when comparing “exemplary” Jigsaw classes with control classes.

In the present study, Jigsaw was conducted by fifth-grade classroom teachers over a school year. The process evaluation assessed classroom implementation of Jigsaw as well as teachers' reactions to the in-service training. The outcome evaluation assessed the effect of Jigsaw participation upon various student attitudes and behaviors. As in our prior study of Jigsaw, schools were matched and randomly assigned to condition. However, in the prior study, recruitment of the eligible teachers for Jigsaw training was poor. In the present study, we were able to use the recommendations of local teachers to improve substantially teacher recruitment, thus reducing the threat of a selection bias by increasing the likelihood that Jigsaw and control classes were initially equivalent. Based upon our prior process evaluation results, the Jigsaw in-service training was revised in an attempt to boost the frequency and quality of Jigsaw implementation. Finally, a hierarchical data analysis strategy was followed that did not violate assumptions about statistical independence.

**METHOD**

**Design**

Jigsaw was offered during the third year of a 3-year longitudinal study. In fall 1978 (when the study began), the students were third graders in 13 elementary schools (grades K–6) in a public school district in Northern California. Schools were paired based on characteristics of their students, faculties, principals, and special programs, and one school from each pair was randomly assigned to the experimental condition and the other to the control condition. The details of this procedure have been described earlier (Moskowitz, Schaps, & Malvin, 1980). During the first 2 years of the study, the cohort of experimental students was exposed to teachers who were trained in Magic Circle; however, Magic Circle was found to have had no systematic effects (Moskowitz, Schaps, & Malvin, 1982; Schaeffer, Moskowitz, Malvin, Schaps, & Condon, 1981).

At the beginning of the third year, the teachers of the experimental cohort were offered Jigsaw in-service training. Eleven of the 13 fifth-grade teachers in the experimental schools participated in the in-service training and conducted Jigsaw in their classrooms. Comparing all experimentals with all controls would provide an insensitive test of Jigsaw's effectiveness, because not all of the experimental students were exposed to Jigsaw. Hence, students whose teachers completed Jigsaw training (participants) were compared with students in the control schools (controls). The present study employed a nonequivalent control group.

---

1 Because 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.
design, with a pretest at the end of the second year and a post-test at the end of the third year.

Subjects

The subjects were fifth-grade students during the 1980–1981 school year. In fall 1980, there were 231 students enrolled in the 11 participant classes, and 249 students were enrolled in the 13 control classes. However, 48 (21%) of the participants and 48 (19%) of the controls were excluded, due to attrition or lack of parental permission for testing. The participant group consisted of 89 boys and 94 girls. The control group consisted of 106 boys and 95 girls. The ethnic composition was 88% white, with Mexican-Americans (5%) comprising the largest minority group.

In-Service Training Program

Two-hour in-service training sessions were held once a week for 10 weeks. There was an additional 2-h review session 11 weeks after the training ended. During and after the training, the trainer assisted teachers in their classrooms. Teachers who completed the in-service training were paid $200 and were offered postgraduate credit.

Measures

Process evaluation data. Several methods were used to monitor implementation of Jigsaw, and to ascertain reactions to the Jigsaw training: (a) surveys of teachers after each training session, at the end of training, and at the end of the school year; (b) classroom observations of Jigsaw by a trained observer; and (c) weekly reports from teachers on how much class time was devoted to Jigsaw.

Student self-report outcome data. Student self-report data were collected with the Student Questionnaire. The pretest was administered in May 1980 and the post-test in May 1981. The Student Questionnaire assesses cooperative and competitive classroom climate (Cooperative Climate and Competitive Climate), affective teaching climate (Affective Climate), attitudes toward school (Attitude School), academic and social self-esteem (Academic Self and Social Self), attitudes toward peers (Attitude Peers), and locus of control for success and failure (Control Success and Control Failure). The reliabilities (coefficient α) for these scales ranged from .56 to .91 (median = .73).3

Student archival outcome data. The measures of achievement were the total reading (Read) and mathematics (Math) stanine scores from the Stanford Achievement Test. The measure of student attendance (Attend) was the average monthly number of absences from January through April. These data were obtained for the spring of 1980 and 1981.

Teacher reports on student behavior. Teacher ratings of student misbehavior were obtained at pretest and post-test with the Student Behavior Report.4 Using the class roster, each teacher indicated how frequently each child had been a minor (Minor) and major (Major) discipline problem during the previous 4 months (January through April). A 5-point scale was used that ranged from “never” to “about once a day or more.”

Data Analysis

With a nonequivalent control group design, differences between the groups at post-test may be due to selection biases and not due to the treatment. Thus, pretest data were

2 See Tuck (1981) for the training curriculum.
3 See Moskowitz, Condon, Brewer, Schaps, and Malvin (1979) for details of the scaling.
4 The teachers who provided the pretest data were the students’ teachers in the prior year.
subjected to analyses of variance to explore potential biases due to initial nonequivalence. Post-test data were subjected to analyses of covariance with the corresponding pretest as the covariate, to control for some preexisting differences.

A hierarchical analysis strategy was utilized, to avoid violating assumptions about the statistical independence of observations (Hopkins, 1982). Students were nested within class; classes were nested within school; schools were nested within condition; and sex was crossed with condition, school, and class.\(^5\) Condition and sex were treated as fixed factors; and schools, classes, and students were treated as random factors. Hence, the school term was used in testing the condition main effect and the school \(\times\) sex term was used in testing the condition \(\times\) sex effect. Preliminary analyses were conducted to determine whether the school term could be pooled with the class term and the class term with the residual; or if the school \(\times\) sex term could be pooled with the class \(\times\) sex term and the class \(\times\) sex term with the residual. Pooling these terms increases statistical power in the tests of the condition and condition \(\times\) sex effects (the two texts that are of primary interest, the results of which are summarized in the subsequent section). As recommended by Winer (1971), Type I error was set at .25 for these preliminary analyses, in order to minimize Type II error. For all other analyses, Type I error was set at .05.

RESULTS

Process Evaluation

The process data indicated that teachers in the participant group found Jigsaw 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. The amount of time that teachers devoted to Jigsaw ranged from 1 to 4 h per week over a 24-week period, and averaged almost 2 h per week \((M = 1.91, SD = .87)\). The observations indicated that Jigsaw was implemented in eight of the participant classes. However, in three of these classes a substantial amount of student off-task behavior was observed. Hence, there were only five "exemplary" classes in which Jigsaw was implemented proficiently; these classes devoted an average of 2.24 h per week to Jigsaw \((SD = 1.11)\). In the remaining three classes, either no student teaching was observed, or the curriculum had not been "jigsawed"; in these classes, students worked in small groups on a common assignment.

Primary Analysis

Analysis of variance was performed on each of the 14 pretest measures to test for the initial equivalence of participant and control conditions. A significant main effect for condition was found on Attitude Peers, \(F(1,11) = 6.32, p < .05\). As compared to control students, Jigsaw participants initially had less positive attitudes toward their peers. No significant condition \(\times\) sex interactions were obtained.

The results of covariance analyses of the post-test data are summarized.

\(^5\) There were too few minority students to use ethnicity as a factor in the design. The results were unaffected by excluding these students from the analyses.
TABLE 1
SUMMARY OF POST-TEST RESULTS

<table>
<thead>
<tr>
<th>Measure</th>
<th>Participants vs controls</th>
<th>Exemplary participants vs controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Condition</td>
<td>Condition x sex</td>
</tr>
<tr>
<td></td>
<td>$F(1,10)$</td>
<td>$F(1,10)$</td>
</tr>
<tr>
<td>Cooperative Climate</td>
<td>$&lt;1$</td>
<td>$&lt;1$</td>
</tr>
<tr>
<td>Competitive Climate</td>
<td>$&lt;1$</td>
<td>$&lt;1$</td>
</tr>
<tr>
<td>Affective Climate</td>
<td>$&lt;1$</td>
<td>$1.22$</td>
</tr>
<tr>
<td>Attitude School</td>
<td>$&lt;1$</td>
<td>$&lt;1$</td>
</tr>
<tr>
<td>Academic Self</td>
<td>1.00</td>
<td>5.78*</td>
</tr>
<tr>
<td>Social Self</td>
<td>$&lt;1$</td>
<td>1.71</td>
</tr>
<tr>
<td>Attitude Peers</td>
<td>$&lt;1$</td>
<td>$&lt;1$</td>
</tr>
<tr>
<td>Control Success</td>
<td>1.18</td>
<td>$&lt;1$</td>
</tr>
<tr>
<td>Control Failure</td>
<td>$&lt;1$</td>
<td>$&lt;1$</td>
</tr>
<tr>
<td>Attendance</td>
<td>$&lt;1$</td>
<td>$&lt;1$</td>
</tr>
<tr>
<td>Read</td>
<td>$&lt;1$</td>
<td>2.25</td>
</tr>
<tr>
<td>Math</td>
<td>5.24*</td>
<td>1.18</td>
</tr>
<tr>
<td>Minor</td>
<td>2.14</td>
<td>$&lt;1$</td>
</tr>
<tr>
<td>Major</td>
<td>$&lt;1$</td>
<td>$&lt;1$</td>
</tr>
</tbody>
</table>

* $p < .05$.
** $p < .01$.

in Table 1 (left portion). The reported results utilized school and school x sex for error terms, as pooling error terms did not affect the interpretation of the results. Only two significant effects were obtained: a main effect on Math (which reflected lower mathematics achievement among Jigsaw participants), and a condition x sex interaction on Academic Self. Analysis of the simple effects indicated that female participants had greater academic self-esteem than female controls, $F(1,10) = 6.46, p < .05$; however, male participants did not differ from their controls, $F(1,10) < 1$.

Exemplary Jigsaw Analysis

The primary analysis did not yield a pattern of treatment effects. These results could be due to variation in the frequency and quality of Jigsaw implementation. A second analysis was performed to determine whether a pattern of effects would be found in the five exemplary Jigsaw classes.

To test for initial equivalence, students from the exemplary Jigsaw classes were contrasted with control students at pretest. The only significant condition-related effect obtained was a condition x sex interaction on Control Failure, $F(1,9) = 5.33, p < .05$. However, the simple effects

---

6 For many variables, pooling the school term with the class term, and the school x sex term with the class x sex term was permissible, but had little effect on the tests of condition and condition x sex. Pooling the class terms with the residual was rarely warranted, indicating that students within each class were interdependent.
for condition were not significant for either boys, $F(1,9) = 1.67$, or girls, $F(1,9) = 3.47$; thus, there was little evidence for initial nonequivalence.

The post-test results are summarized in Table 1 (right portion). Significant condition $\times$ sex interactions were obtained on Academic Self and Social Self. For Academic Self, neither the condition effect for boys, $F(1,8) = 2.35$, nor for girls, $F(1,8) = 3.24$, was significant. For Social Self, the condition effect was significant for boys, $F(1,9) = 5.31$, $p < .05$, but not for girls, $F(1,8) = 1.62$. Boys in exemplary Jigsaw classes had lower social self-esteem than did their controls.

**DISCUSSION**

Participation in Jigsaw did not have any positive effect on students. Jigsaw failed to influence students' perceptions of classroom climate, attitudes toward peers or school, locus of control, school attendance, or reading achievement. As compared to controls, Jigsaw participants had lower mathematics achievement, and boys in exemplary Jigsaw classes had lower social self-esteem. Although a positive effect was found on academic self-esteem for female participants, this effect was not obtained when girls in the exemplary Jigsaw classes were contrasted with girls in the control classes. The absence of any pattern of effects suggests that the few differences obtained between the participant and control groups were spurious.

Jigsaw's ineffectiveness does not seem to be attributable to methodological weaknesses. Almost all of the eligible teachers in the experimental schools completed the Jigsaw training. In addition, there was little evidence of initial nonequivalence between Jigsaw students and control students. Thus, the present study appears to have utilized an appropriate control group. Furthermore, the analysis strategy provided a sensitive test of the condition-related effects without violating assumptions about statistical independence.

The present results replicate the findings of our prior process and outcome evaluation of Jigsaw (Moskowitz et al., 1983). Across the two studies, the frequency and quality of Jigsaw implementation and teachers' reactions to the in-service training were similar. Neither study found any pattern of effects for Jigsaw participation.

An obvious explanation for our failure to find positive effects in two evaluations of Jigsaw would be poor implementation of the technique. However, in our two studies, as compared to other studies, weekly exposure to Jigsaw was fairly typical, and overall exposure to Jigsaw was much higher. In our studies, Jigsaw participation averaged about 2 h per week over a 24-week period, yielding approximately 48 h of total exposure. Other studies had 2 to 4 h per week of Jigsaw participation, but they lasted only 2 to 8 weeks, yielding 8 to 24 h of overall exposure.
(Blaney et al., 1977; Geffner, undated; Lucker et al., 1976). Unfortunately, other studies did not assess the quality of Jigsaw implementation; thus, it is impossible to contrast studies on this dimension. Taking the originators' description as our guide to appropriate Jigsaw implementation (Aronson et al., 1978), our observations revealed considerable variation among Jigsaw classes. In our two studies, only 8 of the 19 Jigsaw classes implemented the technique proficiently. Contrasting students in these exemplary Jigsaw classes with students in control classes did not reveal positive effects for Jigsaw participation. Thus, implementation quality does not seem to be an adequate explanation for Jigsaw's ineffectiveness.

Given our present results and the lack of convincing evidence from prior research, Jigsaw does not appear to be a useful strategy for producing affective benefits (Geffner, undated; Lucker et al., 1976; Moskowitz et al., 1983). Yet various other cooperative learning strategies have had positive influences in the affective domain (Johnson, 1980; Sharan, 1980; Slavin, 1980). There is, however, an important theoretical difference between Jigsaw and other cooperative classroom strategies. According to Slavin (1980), cooperative learning "is primarily a change in the interpersonal reward structure of the classroom." However, unlike most other cooperative learning techniques, with Jigsaw there is no group product, nor do students receive grades based upon their group test performance. In the Jigsaw classroom, like the traditional classroom, the reward structure is individualistic or competitive. Jigsaw's originators rejected a cooperative reward system because they believed it would generate student resentment and parental complaints (Aronson et al., 1978). Hence, they sacrificed a major theoretical component of cooperative learning for pragmatic reasons.

Other researchers have stressed the importance of a cooperative reward structure for promoting helping behavior and academic performance. In a recent review of the research on student interaction and learning in small groups, Webb (1982) concluded, "Rewarding students for the achievement of all group members consistently promoted helping behavior. Instructing students to work with others was not always effective unless accompanied by group rewards." Johnson and Johnson (1974) have argued that "The motivation of students to do well on final examinations, furthermore, would operate to minimize the effects of the way in which discussion groups were structured prior to the examination."

We suspect that Jigsaw's failure to produce affective benefits is due to its reward structure. Jigsaw could be modified to utilize a cooperative reward structure by assigning grades to students based upon the average test performance of their Jigsaw group. Research on this revised technique should evaluate the implementation process as well as the effect on student outcomes.
REFERENCES


GEFFNER, R. (no date). Effects of cooperative learning on students' attitudes and self-esteem. University of Texas, Tyler.


