

Academic Performance, Prejudice, and the Jigsaw Classroom: New Pieces to the Puzzle

IAIN WALKER* and MARY CROGAN

Psychology Division, Murdoch University, Murdoch, WA 6150, Australia

ABSTRACT

This field study investigated the effects of a cooperative learning environment and a Jigsaw classroom environment on academic performance, self-esteem, liking of school, liking of peers, and racial prejudice. The subjects were 103 children in Grades 4–6, in two separate schools. The cooperative learning condition was used as a baseline measure of the effects of cooperation, against which the effects of a Jigsaw method, involving both cooperation and interdependence, were compared. The results reveal that Jigsaw produced significant improvements on measures of academic performance, liking of peers, and racial prejudice. In contrast, the effect of the cooperative condition was to exacerbate pre-existing intergroup tensions. The present findings demonstrate that the Jigsaw method can be applied successfully in Australian conditions, and lend support to Allport's contact hypothesis. © 1998 John Wiley & Sons, Ltd.

Key words: Jigsaw classroom; prejudice; interdependence; cooperative learning

The contact hypothesis claims that 'Prejudice ... may be reduced by equal status contact between majority and minority groups in the pursuit of common goals ... sanctioned by institutional supports ... provided it is of a sort that leads to the perception of common interests and common humanity between members of the two groups' (Allport, 1954, p.281). Intergroup contact under the wrong conditions, though, can increase conflict (Brown, 1995; Pettigrew, in press).

Considerable research has focused on intergroup relations in the classroom, generally producing mixed and disappointing results (Stephan, 1985). Students' academic performance, self-esteem, attitudes to school, and interpersonal and intergroup friendships are all affected by the method of learning applied in classrooms (e.g. Weigel, Wisner and Cook, 1975). Several groups of researchers have examined the effects of cooperative learning structures in the classroom on students' academic performance, self-esteem, and attitudes toward school and peers. These groups differ in the relative emphasis on group vs. individual reward structures and in their reliance on different kinds of interdependence to produce cooperation (e.g. the Student Team Learning approach developed by Slavin (1989, 1990) and his colleagues; the Circles of

* Correspondence to: I. Walker, Psychology Division, Murdoch University, Murdoch, WA 6150, Australia.

Learning method developed by Johnson and Johnson (1975), Johnson *et al.* (1984); the Jigsaw method of Aronson and his colleagues (Aronson and Patnoe, 1997); and the Groups Investigation technique of Sharan and Sharan (1976). Reviews of these different approaches to cooperative learning (e.g. Miller and Davidson-Podgorny, 1987; Slavin, 1990) point to the general effectiveness of cooperative learning techniques (relative to traditional classroom structures) in producing increases in achievement and in various attitudinal domains. However, the mechanisms responsible for those gains remain unclear. The present study focuses on the Jigsaw technique and its effect on intergroup attitudes.

The *Jigsaw classroom* (Aronson *et al.*, 1978; Aronson and Patnoe, 1997) exploits both cooperation and interdependence in its attempt to maximize intergroup harmony and educational gain. Students in a classroom are arranged into groups which are evenly balanced according to race, ethnicity, sex and academic ability. Each student in each Jigsaw group must learn a unique segment of information, which he/she then teaches to the other members of the Jigsaw group. The Jigsaw group members therefore depend on one another to acquire the composite parts, which, as they are combined, constitute the entire lesson. Because of the structure of the situation, the students have equal status contact (each has a unique and necessary piece of information), they work interdependently (each depends on the others to be able to achieve their desired goals), and they work in pursuit of a common goal (good grades, learning, teacher praise), all with the sanction of authorities (teachers). The Jigsaw method, therefore, provides a classroom structure which meets Allport's (1954) conditions for optimal intergroup contact. Jigsaw is not just a cooperative learning technique:

It is the element of 'required' interdependence among students which makes this a unique learning method, and it is this interdependence that encourages the students to take an active part in their learning (Aronson *et al.*, 1978, p. 28; emphasis in original).

Several studies document Jigsaw's effects (e.g. Araragi, 1983; Aronson *et al.*, 1978; Aronson and Bridgeman, 1979; Blaney *et al.*, 1977), generally showing positive effects on academic performance, liking for school and liking for peers. Questions remain, though, about why the Jigsaw method produces these effects. Aronson and Osherow (1980) claim that increased liking for a student is fundamental to that child's acceptance, which in turn is a prerequisite for the development of positive self-esteem. They further assert that negative self-esteem may adversely affect academic performance. Aronson and Bridgeman (1979) claim that positive effects on academic performance are due to students' increased participation in learning, combined with reduced anxiety. They claim that empathic role-taking increases and that individuals' attributional patterns for success and failure change.

Although Aronson *et al.* (1978, p. 28) claim that *interdependence* is the essential element in Jigsaw, none of the putative mechanisms for Jigsaw's success pertain only to interdependence. All can be seen in solely cooperative classroom structures. Furthermore, recent studies of the effects of cooperation on relations between group members in artificial, laboratory-based groups show that cooperation alone can have at least some of the effects claimed to be due in the Jigsaw method to interdependence (e.g. Desforges *et al.*, 1991; Gaertner *et al.*, 1990).

The Jigsaw technique is *both* cooperative and interdependent. We can therefore ask whether Jigsaw works because it fosters cooperative relations among group members

or because it fosters interdependence. This provides the primary aim of the present report. We note parenthetically that this study was not designed with this aim in mind. Rather, through the vagaries of field research, it changed to become a test of the relative effects of a cooperative learning condition and a Jigsaw (cooperative learning plus interdependence) condition.

There are several forms of interdependence (Miller and Davidson-Podgorny, 1987). In *task* interdependence, each member of a group must contribute to the group's task, otherwise the task cannot be completed successfully. Task interdependence can imply *resource* interdependence, in which group members must rely on others for provision of resources needed to complete the task; and/or it can require *cooperation* for the successful completion of the task. In *fate* interdependence, group members all share the same fate because of their membership in a particular group. This can refer to a *positive outcome interdependence*, in which each member's outcome is a direct function of other group members, or to a *negative outcome interdependence*, in which provision of a particular outcome to one member precludes the same outcome being provided to another member (i.e. direct competition for scarce resources). The Jigsaw method is based on task interdependence, because group members each possess a unique piece of information (a valued resource) necessary for the group to complete the task successfully. However, in usual Jigsaw implementations, students' performances are tested and rewarded individually, so there is normally no fate interdependence.

This study was originally designed to see if the Jigsaw method would work in the social climate of Australia. The grave social problems facing Aborigines (Callan, 1986) and the overwhelmingly negative content of the stereotype of Aborigines (Walker, 1994) led us to wonder whether typical Jigsaw results would be attained in classrooms in Australia. Also, hostility and resistance to recent increases in Asian immigration (Ho, 1990), problems of language, and of resentment at perceived 'over-achievement' by Asians, led us to suspect that the Jigsaw may not have the same effects in Australia as in the USA.

The Jigsaw classroom has, to our knowledge, only once been implemented in Australia, and that study (Macrae *et al.*, 1977) focused on university psychology students' acquisition of statistics skills under Jigsaw and normal teaching routines. The original aim of this study, then, was simply to explore the efficacy of the Jigsaw method in a different social climate. The vicissitudes of stepping outside the laboratory led this aim to be perverted over the course of the study. While we still managed to explore the applicability of the Jigsaw classroom, almost by accident the study evolved to be, in addition, a test of the relative contributions of cooperation and interdependence to the effects of the Jigsaw classroom.

METHOD

Participants

A rural private school agreed to participate in the study. Because of the distance between us and this town (1,600 km), the program was organized over the telephone. The sixth-grade class was to implement the program and the fifth-grade class was to serve as the control. Unfortunately, we could not randomly assign classes to conditions, assignment being determined by the teachers' cooperativeness and by the ages of children (see Aronson *et al.*, 1978, for issues of children's ages in Jigsaw

classrooms). The two classes had similar proportions of Asians and European-Australians, but the control class had no Aboriginal students.

Although the teacher of the experimental class initially agreed to implement the Jigsaw method, after two weeks of familiarizing students with the group structure and intragroup cooperation, she modified the method so that students worked cooperatively in their groups to complete the syllabus. This was done without consulting us, and we only learned of it much later. The teacher made the changes to the program for understandable and pragmatic reasons—the large class size and the disruptiveness of some students. After the project had been running for four weeks, the school lost a teacher—the fifth-grade control group teacher—and the entire school was reorganized. We decided to collect the post-test data, and terminate the program.

A second school—a state school in metropolitan Perth—agreed to participate because of Jigsaw's potential to foster language development and cross-cultural interaction. An authentic and complete intensive three-week Jigsaw program was run in a fourth-grade class, and a split fourth/fifth-grade class served as the control. Again, we could not randomly assign classes to conditions.

At the end of the study, we had data from four classes in two schools. In the first school (school A), a cooperative learning program had been implemented for four weeks in a sixth-grade class, and a fifth-grade class served as a control group. In the second school (school B), a genuine Jigsaw program was run for three weeks in a fourth-grade class, and a fourth/fifth-grade class served as a control group. Thus, we ended with two programs, each with a same-school control.

Subjects

At school A, there were 31 students in the experimental group and 29 students in the control group. At school B, the numbers were 20 and 23, respectively.

Procedure

Teacher preparation. The teachers of the experimental classes at schools A and B read a description of the Jigsaw program (Aronson and Goode, 1980), and the key facets of the program were discussed in detail—in person with the teacher at school B and on the telephone with the teacher at School A.

Groups. The students in the two experimental classes were divided into Jigsaw groups by their teachers, so that ethnicity, academic ability, and sex were distributed as evenly as possible within and across groups. No group contained best friends or worst enemies. The class at school A was divided into four groups of six students, and the class at school B into four groups of four. These sizes were chosen by the teachers on the grounds of class size and available space for groupwork. Each class had 'extras' because the class sizes were not evenly divisible by the group sizes. In both classes, these students were deliberately chosen because they were the lowest achievers in their class. They were paired with high achievers who were 'good helpers' and who served as role models in the group activities.

Student preparation. Prior to the program, students in each experimental group familiarized themselves with their group peers, practised their new roles as peer tutors, and practised particular relevant skills such as reading for meaning, discussing main ideas, listening, and quizzing peers on important information.

During the fortnight before starting the Jigsaw method, the teacher of the experimental group in school A implemented trial runs of the program. It was during the second week that the teacher decided to modify the method from Jigsaw to cooperative learning. The students in the experimental class in school B were used to sitting in groups and only required one week to become familiar with the Jigsaw method prior to implementing the program.

Intensity. The experimental group in school A received a cooperative learning program for 90 minutes each day, twice a week, for four weeks. The experimental group in school B received a Jigsaw program for one hour each day, five days a week, for three weeks.

Curriculum. The implementation of the cooperative method varied slightly from lesson to lesson. The range of activities included collaboratively constructing food charts, sharing information to compile a joint list of nutrient functions, jointly answering review questions, and jointly constructing and solving word games and puzzles involving the key health words.

The teacher in the Jigsaw group divided the lesson topic into four parts. Each group member was given one part. Students with the same segment met in an 'expert group' to read, identify and discuss their key information for 15 minutes. They then returned to their Jigsaw groups, and each person, in turn, taught their unique piece of information to the group. After this, each student asked questions of the others, to revise the key points. After approximately 20 minutes the students were given a printed sheet containing multiple choice questions. To encourage cooperation rather than competition, the students were praised for their cooperation. Whenever anyone achieved full marks on the test, the whole group was praised for being good teachers, and the particular student was praised for being a good listener.

Several children in the experimental group in school B were immigrants, and were only acquiring, or had only just acquired, English skills. This posed problems, which we attempted to overcome by underlining the main ideas in the passages which were being learned during the first week of the program. In the second week, only the key words were underlined. In the final two weeks, the students identified the key teaching points themselves.

Measures

All students were given a pre-test and a post-test battery of measures. All teachers agreed to provide academic performance data for their students, but neither the two control group teachers nor the experimental group teacher in school A provided these data. Consequently, we only have academic performance data for children in the Jigsaw group.

Self-esteem. The Piers–Harris Children's Self-concept Scale (CSCS) was used to measure self-esteem (Piers, 1984). The CSCS is an 80-item scale with good internal consistency, temporal reliability and content validity, designed specifically for use with children and adolescents, and is appropriate for use in Australia (Center and Ward, 1986).

Sociometric class survey. All students rated each of their classmates according to how much they would like to work with, and how much they would like to play with, him/her (1 = a little, 5 = a lot). Ratings of this sort are stable and correlate highly with peer nomination measures (Gronlund, 1959). The two control groups were used to assess reliability. In school A, where the measures were administered eight weeks apart, the correlation between pre-test and post-test scores was 0.66 for the 'work with' ratings and 0.57 for the 'play with' ratings (both have $p < 0.001$). In school B, the respective correlations were 0.67 and 0.54 (again, both have $p < 0.001$). The sociometric ratings are thus temporally stable, at least in groups where we expect no change.

Racial prejudice measures. These measures assessed students' attitudes to Asians, Aborigines and European-Australians. Two measures were used, one of social distance and one of stereotypes. On each measure, low scores indicate more prejudice.

The social distance measure presented six photographs, unambiguously depicting one girl and one boy (aged 8–11) each of Aboriginal, Asian and European-Australian background. The photographs were matched in facial expression, hairstyle, dress and general appearance. For each photograph, subjects rated (from 1 = not at all, to 4 = very much) how much they would like to (a) play team sport with this person, (b) play with this person at lunchtime, (c) play with this person on the weekend, and (d) have this person to their birthday party. Scores were summed across the four situations, producing an index of social distance with high scores representing less social distance.

The second measure of racial prejudice required subjects to indicate how much they thought each of seven adjectives (honest, friendly, clever, hard-working, kind, clean, and helpful) described the three groups Aborigines, Asians, and European-Australians (1 = no, 2 = sometimes, 3 = most of the time, 4 = yes). Responses were summed across the seven adjectives to give a general evaluation index. Although the positive wording of the adjectives threatens measurement validity, it was a necessary concession to concerns expressed by teachers in school A that completing the scale might increase students' prejudices.

The two measures of prejudice produced reasonable test–retest correlations for the two control groups (median $r = 0.60$), indicating good temporal stability.

RESULTS

Academic performance

Academic performance data came only from the Jigsaw group at school B, rendering the test of the hypothesis inconclusive. Nonetheless, the group's results are presented, since they suggest an important effect.

The pre-test and post-test scores were obtained by averaging each student's daily test scores from the first five days and the last five days of the program, respectively. The mean difference was significant (pre-test $M = 8.73$ and $SD = 1.59$; post-test $M = 9.95$ and $SD = 1.31$; $t = 4.95$, one-tailed $p < 0.0005$), representing a performance gain of almost one standard deviation across the four-week program. A two-way (sex \times race) analysis of variance (ANOVA) on the difference scores produced no significant main effects or interaction. Hence, gains appeared to be evenly spread across groups.

Table 1. Mean changes in self-esteem

	School A		School B	
	Experimental	Control	Experimental	Control
Asian girls	5.80	1.67	4.44	0.77
Asian boys	-3.33	3.67	5.20	2.75
Euro-Australian girls	0.80	2.50	2.00	-5.00
Euro-Australian boys	4.20	-1.00	1.25	3.50

Self-esteem and liking for school

A $2 \times 2 \times 2 \times 2$ ANOVA (condition \times school \times race \times sex) of gain scores on the self-esteem and liking for school measures was performed. For self-esteem, the four-way interaction was almost significant ($F(1,68) = 3.00$, $p = 0.088$). Table 1 summarizes the mean gain scores for Asian girls and boys and for European-Australian girls and boys, by school and by condition. No other main effect or interaction were significant.

The main effect of condition on liking for school scores was almost significant ($F(1,84) = 3.55$, $p = 0.063$). The mean gain scores in the two experimental conditions were 2.55 and 1.55 in schools A and B respectively, and in the two control conditions they were 0.86 and -0.95, respectively. No other effects approached significance.

Liking of peers

Four-way ANOVAs (school \times condition \times race \times sex) were calculated for the gain scores on the 'working with' and 'playing with' variables separately. There was a significant two-way interaction between school and condition ($F(1,78) = 10.40$, $p = 0.002$) on work with ratings. Tukey's HSD test (Kirk, 1990) revealed that the experimental (0.16) and control (-0.26) group means in school B differed significantly ($q = 3.97$, $df = 80$, $p < 0.05$). The difference between the two means in school A (-0.05 and 0.34 in the experimental and control groups, respectively) was not significant. There was also a significant four-way interaction ($F(1,78) = 4.58$, $p = 0.035$) which proved impossible to interpret.

The school \times condition interaction on 'play with' ratings was almost significant ($F(1,78) = 3.45$, $p = 0.067$). Ratings decreased in the school A experimental (-0.17) and school B control (-0.20) conditions, and increased in the school A control (0.25) and school B (0.04) conditions.

Liking for ingroup and outgroup peers

The gain scores of peers' work with and play with ratings of students, representing separately the Asians and the European-Australians within each classroom, were analysed by a four-way ANOVA (school \times condition \times race \times sex). As Aborigines were present only in school A's experimental class, one-way ANOVAs were performed on the data representing the three racial groups within that class.

There was a significant school \times condition interaction ($F(1,78) = 10.21$, $p = 0.002$) on students' work with ratings of Asian peers. In school A, ratings decreased in the

control condition (-0.09) and increased in the experimental condition (0.19). In school B, the experimental group's ratings increased (0.39) and those of the control group decreased (-0.22). The difference between the two conditions in school B was significant ($q = 3.88$, $df = 60$, $p < 0.05$).

The same two-way interaction appeared on work with ratings of European-Australian peers ($F(1,78) = 7.55$, $p = 0.007$). At school A, the experimental group decreased (-0.08) and the control group increased (0.15). At school B, the experimental condition increased (0.22) and the control condition decreased (-0.42). The difference between the last two groups was significant ($q = 3.70$, $df = 60$, $p < 0.05$). There was also a significant, and difficult to interpret, four-way interaction ($F(1,78) = 6.04$, $p = 0.016$) on this variable.

ANOVAs on the work with ratings of Aboriginal peers, and on Aborigines' ratings of Asian and European-Australian peers in the experimental class in school A, produced no significant main effects or interactions.

There was only one significant effect on 'play with' ratings of Aboriginal peers: a two-way interaction between sex and race ($F(1,78) = 8.73$, $p = 0.016$). In school A, Aboriginal girls increased their ratings of Aboriginal peers (0.41), while Aboriginal boys decreased their ratings (-0.94). Asian girls' ratings of Aboriginal peers decreased (-0.63), while Asian boys' ratings increased (0.50). The difference between Aboriginal girls' and boys' ratings of Aboriginal peers was significant ($q = 3.23$, $df = 9$, $p < 0.05$).

Social distance ratings

In the social distance measures, each target group was represented by a picture of a boy and a girl. Ratings of these two pictures were highly correlated within each target group, and so were summed. Gain scores were analysed using four-way ANOVAs (school \times condition \times sex \times race), separately for each target group.

There was a two-way interaction between school and condition ($F(1,62) = 6.44$, $p = 0.014$) on ratings of Asian children. Scores for the school A experimental condition decreased (-0.23) while those in the control condition increased (1.16). This difference was significant ($q = 4.20$, $df = 60$, $p < 0.01$). Both conditions in school B increased (1.67 and 0.65 for the experimental and control groups, respectively). The difference between these two was significant ($q = 4.40$, $df = 60$, $p < 0.05$).

The ratings provided by the Aboriginal students in the experimental group in school A were analysed using paired *t*-tests, and produced no significant differences.

There was a three-way interaction between school, condition, and race ($F(1,62) = 7.44$, $p = 0.008$) on ratings of Aboriginal children. European-Australian students' ratings of Aboriginal children in the school A experimental group decreased (-0.32), but increased in that school's control group (0.22). In the school B experimental group, European-Australian students' ratings of Aborigines decreased (-0.20), while in the control group they increased (1.00). This difference was significant ($q = 4.29$, $df = 60$, $p < 0.05$). Asian students' ratings remained the same in the school A experimental condition, and increased in that school's control group (0.64). In school B, Asian students' ratings increased in both the experimental (0.68), and control (0.06) groups.

There was a school \times condition interaction ($F(1,62) = 9.91$, $p = 0.003$) on social distance ratings of European-Australian children. Students' ratings increased in the

school B experimental (2.12) and control (0.26) conditions, decreased in school A's experimental condition (-0.54), and increased in school A's control condition (1.09). The difference between the two conditions in school B was significant ($q = 5.87$, $df = 60$, $p < 0.01$), as was the difference between the two experimental conditions ($q = 5.87$, $df = 60$, $p < 0.01$).

Stereotype ratings

Each student's pre and post-ratings of Asians, Aborigines and European-Australians on seven adjectives were summed. The gain scores were then analysed separately by race of target using a four-way ANOVA (school \times condition \times race \times sex). The differences between the pre- and post-test ratings made by the Aboriginal students in the school A experimental group were evaluated using paired t -tests. None of the differences was significant.

Ratings of Asians. There was a school \times condition interaction ($F(1,50) = 8.84$, $p = 0.005$) on ratings of Asians. The mean rating increased in the school A experimental (0.21) and control (0.57) conditions, and in school B's experimental condition (0.75), but decreased in that school's control group (-0.20). The difference between the two conditions in school B was significant ($q = 4.77$, $p < 0.01$).

There was also a race \times school interaction ($F(1,50) = 11.75$, $p = 0.001$) on ratings of Asians. Ratings of Asians by Asian students increased in school A (0.58) but decreased in school B (-0.11). European-Australian students' ratings of Asians increased in both school A (0.29) and B (1.11). The difference between the two groups' ratings of Asians in school B was significant ($q = 5.65$, $p < 0.01$).

Ratings of Aborigines. The sex \times school interaction on ratings of Aborigines was almost significant ($F(1,62) = 3.35$, $p = 0.073$). Girls' ratings decreased in the school A experimental condition (-0.10), and increased in the control condition (0.10). In school B, girls' ratings increased in the experimental group (0.60), but decreased in the control group (-0.57). Boys' ratings decreased in the school A experimental condition (-0.13) and increased in the control group (0.29). Their ratings decreased in both the school B experimental (0.04) and control (0.31) groups.

Ratings of European-Australians. There was a school \times condition interaction ($F(1,50) = 14.29$, $p < 0.001$) on students' ratings of European-Australians. In school B, ratings increased in the experimental condition (0.66) and decreased in the control condition (-0.34), but in school A the experimental class decreased (-0.34) and the control class increased (0.22). The difference between conditions in school B was significant ($q = 4.15$, $df = 60$, $p < 0.01$).

DISCUSSION

We discuss the findings of the present study in terms of the different dependent variables, evaluated in the context of the different school settings and different learning methods. Finally, conclusions are drawn from the present findings, and implications for further research in the field discussed.

Academic performance

Consistent with others' findings (e.g. Araragi, 1983), students in the Jigsaw group improved their academic performance. This result must be treated circumspectly, however, since performance data could not be obtained for students in the other groups. No other unusual skill program was being run concurrently to confound the effects of the Jigsaw program. Practice effects were unlikely, since the program was structured so that an increased skill input and attainment level was required of students during each consecutive week.

This finding is interesting in light of the ethnic diversity of the class, and the lack of English proficiency in many students. The findings of Blaney *et al.* (1977) imply that Jigsaw may be unsuited to the learning needs of English as a Second Language students. The present study suggests the contrary.

Self-esteem and liking for school

Self-esteem increased in the experimental groups in both schools, relative to their respective controls, but these gains were not significant. This may have been due to a ceiling effect. The pretest means in both experimental groups were already equal to, or above, the standardized age means (Piers, 1984). Gains reported in the literature are small as well, however. Similarly with liking for school, the lack of significant results may have been due to ceiling effects and to an expected small effect size.

Liking of peers

Work with ratings of peers. Students in the Jigsaw group increased their work with ratings of peers, relative to the control group. Because the Jigsaw class was small (20) the rotation of students in both expert and Jigsaw groups meant that each student interacted with approximately half of the class members. These students were keen to participate and to take their turn as teacher. However, to maintain the lessons' flow and the students' confidence, both teachers had to assist when requested with problems such as the pronunciation and explanation of new vocabulary.

In contrast, students in the cooperative learning group were not as motivated by the prospect of working cooperatively. The class size was larger (32) and many students required frequent external discipline. The teacher reported that these students saw the group activities as an opportunity to misbehave. Students in this group decreased their work with ratings of peers. These ratings were probably influenced by the negative attitudes and behaviours of some of the students. Other problematic factors were: the program may not have been sufficiently intense to affect friendships; school vacations intervened in the program; and the class had been dismantled and students reallocated to new classes two weeks before the post-test measures.

Play with ratings of peers. Neither experimental group demonstrated significant changes in play with ratings of peers. The factors which were mentioned regarding the work with ratings made by the cooperative learning group students probably also affected their play with ratings. Even in the Jigsaw group, the effects of the daily one-hour Jigsaw lessons for a three-week period may not have been sufficiently intense to overcome the well-defined routines of the school playground. The intensity and

duration of the program may therefore play an important role in any cumulative effects of the program.

Liking of ingroup and outgroup peers

The Jigsaw students', but not the cooperative learning students', work with ratings of both the ethnic groups represented in the classroom increased, indicating that Jigsaw enhanced liking of ingroup and outgroup peers in work-orientated relationships. This supports Bridgeman's (1981) claims that changing roles from peer to tutor diminishes egocentrism, and that children therefore have a greater potential to understand children from different ethnic backgrounds.

Social distance ratings

The Jigsaw students' social distance ratings of both Asian and European-Australian children decreased across the program, but the European-Australians' ratings of Aboriginal children increased. The failure of these children's otherwise positive changes to generalize to Aboriginal children may have been because stereotypes about Aborigines are particularly pernicious. Also, the interdependence of Asians and European-Australians in the Jigsaw program perhaps increased understanding and enhanced role-taking abilities between the two groups. Aborigines were not represented in the class and there were few Aborigines at the school, so the children did not have similar opportunities to engage in interdependent activities with Aboriginal children. Thus, the positive effects observed between the two interacting groups did not generalize, at least for children belonging to the majority group.

In school A, the experimental students' ratings of Asian, Aboriginal and European-Australian targets increased, while the control group's ratings decreased, suggesting that the cooperative learning program negatively affected the students' racial attitudes.

Stereotypes

Students in the Jigsaw group decreased the negative traits they attributed to Asians and European-Australians. Coupled with the social distance effects, the Jigsaw appears to have affected positively students' views of the target groups. An exception to this otherwise positive finding was the increased social distance of European-Australian students to Aborigines. This does, however, indicate that the positive changes were genuine effects of the program and were not due to social desirability.

In general, school A's experimental group showed increased stereotyping while the control group showed decreased stereotyping, suggesting that the cooperative learning condition negatively affected intergroup attitudes. Contact between students from different racial backgrounds can, without Allport's optimal conditions, worsen pre-existing prejudices. The imposition of intergroup interaction in school A's experimental condition appears to have exacerbated pre-existing tensions.

CONCLUSIONS

The present field experiment shows that the Jigsaw method is effective in Australian social conditions in producing positive changes in academic performance, attitudes to

peers, and prejudice. The cooperative learning method produced generally negative results. The important element of the Jigsaw classroom for enhancing intergroup relations appears to be interdependence, not cooperation. This is not to say that cooperation is not important. On the contrary, it is an essential element of the Jigsaw method. But it alone is not enough.

There were some differences between the two schools which qualify these results. Two experienced teachers ran the Jigsaw program in school B, and the class was smaller and younger than the cooperative learning class in school A, which was run by a less experienced teacher who was unassisted. There was evidence of intermittent, overt racial prejudice in the cooperative learning class, and about one-third of the students behaved disruptively on occasions. Also, Jigsaw students worked on 'traditional' curriculum areas, while students in the cooperative learning group worked solely on health topics. These potential confounds notwithstanding, the major difference between the two experimental classes was the learning environment. In school A the conditions of Allport's contact hypothesis were met—the students had equal status in the Jigsaw groups, met with the sanction of authority figures, and worked cooperatively in pursuit of a common goal. In school B, these conditions were violated in one way or another.

It was also found that certain other conditions were necessary for the success of Jigsaw. Teacher competence, organization and enthusiastic support are critical to the program's success. Where classes are large, or are composed of minority students for whom English is a second language, teachers must have adequate support. If this is not addressed, then students in racially or ethnically mixed classes, who would often benefit the most from the Jigsaw program, may not achieve the level of interdependence necessary to achieve positive outcomes. Finally, the length and intensity of the program are important considerations, which must often be decided upon in relation to pre-existing school routines and budgets.

ACKNOWLEDGEMENTS

We wish to express our gratitude to the teachers and students involved in this study, and our thanks to Elaine Chapman for her many useful and insightful comments on an earlier version of the paper.

REFERENCES

- Allport, G. A. (1954) *The Nature of Prejudice*, Addison-Wesley, New York.
- Araragi, C. (1982) 'The effect of the Jigsaw learning method on children's academic performance and learning attitude', *Japanese Journal of Educational Psychology*, **31**, 102–112.
- Aronson, E., Blaney, N., Stephan, C., Sikes, J. and Snapp, M. (1978) *The Jigsaw Classroom*, Sage, Beverly Hills, CA.
- Aronson, E. and Bridgeman, D. (1979) 'Jigsaw groups and the desegregated classroom: in pursuit of common goals', *Personality and Social Psychology Bulletin*, **5**, 438–446.
- Aronson, E. and Goode, E. (1980) 'Training teachers to implement Jigsaw learning', in S. Sharan, C. D. Webb and R. Hertz-Lazarowitz (Eds), *Cooperation in Education*, pp. 47–81, Provo, Utah. Brigham Young Press.
- Aronson, E. and Osherow, N. (1980) 'Prosocial behavior and academic performance: experiments in the classroom', *Applied Social Psychology Annual*, **1**, 163–196.

- Aronson, E. and Patnoe, S. (1997) *The Jigsaw Classroom*, 2nd edn, Longman, New York.
- Blaney, N. T., Stephan, C., Rosenfield, D., Aronson, E. and Sikes, J. (1977) 'Interdependence in the classroom: a field experiment', *Journal of Educational Psychology*, **69**, 139–146.
- Bridgeman, D. (1981) 'Enhanced role-taking through cooperative interdependence: a field study', *Child Development*, **52**, 1231–1238.
- Brown, R. J. (1995) *Prejudice: Its Social Psychology*, Blackwell, Oxford.
- Callan, V. (1986) *Australian Minority Groups*, Harcourt, Brace, Jovanovich, Sydney.
- Center, Y. and Ward, J. (1986) 'A note on the use of two self-esteem inventories with Australian schoolchildren', *Australian Psychologist*, **21**, 473–476.
- Desforges, D. M., Lord, C. G., Ramsey, S. L., Mason, J. A., Van Leeuwen, M. D., West, S. C. and Lepper, M. R. (1991) 'Effects of structured cooperative contact on changing negative attitudes toward stigmatized social groups', *Journal of Personality and Social Psychology*, **60**, 531–544.
- Gaertner, S. L., Mann, J. A., Dovidio, J. F., Murrell, A. J. and Pomare, M. (1990) 'How does cooperation reduce intergroup bias?' *Journal of Personality and Social Psychology*, **59**, 692–704.
- Gronlund, N. E. (1959) *Sociometry in the Classroom*, Harper & Brothers, New York.
- Ho, R. (1990) 'Multiculturalism in Australia: a survey of attitudes', *Human Relations*, **43**, 259–272.
- Johnson, D. W. and Johnson, R. T. (1975) *Learning Together and Alone*, Prentice-Hall, Englewood Cliffs, NJ.
- Johnson, D. W., Johnson, R. T., Johnson-Holubec, E. and Roy, P. (1984) *Circles of Learning—Cooperation in the Classroom*, The Association for Supervision and Curriculum Development, Alexandria, VA.
- Kirk, R. E. (1990) *Statistics: An Introduction*, 3rd edn, Holt, Rinehart & Winston, Sydney.
- Macrae, K., Foddy, M., Fry, C. and Pratt, K. (1997) 'The effects of cooperative learning on attitudes and achievement in psychological statistics'. In press.
- Miller, N. and Davidson-Podgorny, G. (1987) 'Theoretical models of intergroup relations and the use of cooperative teams as an intervention for desegregated settings', in C. Hendrick (Ed.), *Review of Personality and Social Psychology, Vol. 9: Group Perspectives in Psychology*, pp. 205–220, Wiley, New York.
- Piers, E. V. (1984) *Piers-Harris Children's Self-concept Scale: Revised Manual*, Western Psychological Services, Los Angeles.
- Pettigrew, T. F. (in press) 'Intergroup contact theory', *Annual Review of Psychology*.
- Slavin, R. E. (1989) *School and Classroom Organization*, Lawrence Erlbaum, Hillsdale, NJ.
- Slavin, R. E. (1990) *Cooperative Learning: Theory, Research, and Practice*, Allyn & Bacon, Needham, MA.
- Sharan, S. and Sharan, Y. (1976) *Small-group Teaching*, Educational Technology Publications, Princeton, NJ.
- Stephan, W. (1985) 'Intergroup relations', in G. Lindzey and E. Aronson (Eds), *Handbook of Social Psychology*, 3rd edn, Vol. 2, pp. 599–658, Random House, New York.
- Walker, I. (1994) 'Attitudes to minorities: survey evidence of Western Australians' attitudes to Aborigines, Asians, and women', *Australian Journal of Psychology*, **46**, 137–143.
- Weigel, R. H., Wiser, P. L. and Cook, S. W. (1975) 'The impact of cooperative learning experiences on cross-ethnic relations', *Journal of Social Issues*, **31**, 219–244.