GROUPING AND ACHIEVEMENT IN COOPERATIVE LEARNING

John Baer

Abstract. Colleges typically group students homogeneously in classes by means of both admission requirements and course prerequisites, but when professors form cooperative learning groups within classes they generally use heterogeneous grouping. Authors compared heterogeneously and homogeneously grouped cooperative learning groups in six paired classes, taught by the same professor using matching syllabi and assignments. Overall, homogeneously grouped students (who were grouped based on achievement on the first test given in the course) significantly outperformed heterogeneously grouped students on the final examination. High- or average-achieving students particularly benefited from homogenous grouping. Low achievers did equally well in either type of group.

Colleges tend to attract and admit students of similar academic abilities and achievements. Although diversity is an important goal of most colleges in their selection of students, one kind of diversity that is generally not sought by colleges is diversity of academic talent, if one understands that to mean the inclusion of students of all levels of ability. A college education is expensive, no matter who is paying the bill, and college failures often have additional costs beyond the obvious monetary ones, such as undermining the self-esteem—or simply wasting the time—of a young person for whom much evidence suggests that attending a particular college would not be a positive experience. Admitting students with academic records that imply that they would be highly unlikely to succeed at a given college would be neither kind nor wise.

Students also screen colleges, of course, using among other criteria the level of academic rigor expected of students to find a school that will suit them (or perhaps just to find the highest-ranked college that will accept them). These combined screening processes result in a measure of homogeneity among students, at least in terms of levels of academic achievement. The stratification does not end with admission, however, because some college courses have prerequisites that further reduce the heterogeneity of students in a given class.

Despite these powerful homogeneity-inducing forces, many college classes evidence a surprising degree of diversity in the achievement of their students, differences that are not solely due to how seriously different students take their studies, but that reflect very different aptitudes and prior achievements. Although the various screening processes mentioned above increase homogeneity, many colleges nonetheless admit students who demonstrate fairly wide ranges of academic achievement and ability. (In fact, many colleges would like to be more selective and work hard to increase what might be called their "diversity" at the high end of the academic achievement scale.) And even when students at a given college appear rather similar in terms of their average high school GPAs or SAT scores, they may vary widely in terms of what they know about a particular subject.

John Baer is a professor in the School of Education at Rider University, in Lawrenceville, New Jersey.

Given the degree of heterogeneity found in some undergraduate classes, might there be advantages to grouping students in these classes based on some measure of prior achievement in or aptitude for the subject? Sometimes the answer is clearly no. In a course that is primarily delivered via lecture, for example, unless one is going to change the content (in which case one is teaching a different course), it hardly matters how able the student in the next seat might be. The degree of heterogeneity among the students in such a course is not really an issue because every student is being delivered essentially the same course. (What they are "receiving"—what they are attending to and understanding—is another matter, of course, but this is not a function of how students are grouped in the course, because there is no grouping.)

In a class in which students work together frequently and a significant part of students' time in class is devoted to cooperative learning experiences of one kind or another, however, the ways in which students are grouped for instruction might significantly influence how, and how well, students learn. It was the goal of this study to improve our understanding of how such grouping techniques—either grouping students of similar achievement together or grouping students in ways that ensure diversity of prior achievement in every group—might influence academic achievement in a course that employs cooperative learning as a significant instructional component.

Research on Grouping in Cooperative Learning

Cooperative learning is a powerful technique that has been shown to increase student achievement in a wide variety of studies (Hertz-Lazarowitz and Miller 1992; Johnson and Johnson 1989; Slavin 1980, 1984, 1992; Watson 1991). Cooperative learning is most often associated with heterogeneous grouping—as Watson and Marshall (1995) suggest, "heterogeneous grouping of students in cooperative learning is so commonly accepted that it is often included as part of the definition of cooperative learning" (292)—although homogeneous groups are also possible (Lawrence and Munch 1984; Woolfolk 1998).

Grouping students for instruction based on past achievement is a controversial practice (Loveless 1999; Tobias 1989). Although high-achieving students generally achieve more when grouped homogeneously, the outcome for low-achieving students is more varied (Allan 1991; Kulik and Kulik 1982, 1984; Slavin 1990a; Sternberg and Willard 2002). Although both homogeneous and heterogeneous grouping based on prior achievement commonly is used at all grade levels, heterogeneous grouping is more common in elementary schools and homogeneous grouping is more widespread in secondary schools (Loveless 1999). In studies comparing the effects of different grouping practices, academic achievement is typically measured by grades in particular subjects, teacher recommendations, and/or test scores, and the subjects of such studies generally represent a wide (and normally distributed) range of student achievement. Comparisons among three groups (high-, average-, and low-achieving students, based on past performance) are most common in such studies (Allan 1991; Kulik and Kulik 1982, 1984; Slavin 1990a).

Most of the research on the effects of cooperative learning has been conducted in elementary and secondary schools, and the impact of cooperative learning methods is sometimes different when used with undergraduates. As an example of such a difference, the use of group incentives has been found crucial to the success of cooperative learning in elementary schools (Slavin 1984, 1990b, 1991, 1992; Webb 1992), but group incentives do not appear to influence achievement in college-level cooperative learning studies (Baer and Baer 1996; Watson and Marshall 1995).

Very few studies have compared the effects of different methods of grouping students when cooperative learning has been used in college classrooms. Self-selected undergraduate cooperative learning groups have been found to achieve less than teacher-selected groups, whether homogeneously or heterogeneously grouped, in science laboratory situations (Lawrenz and Munch 1984). Similarly, Baer and Baer (1997) found in an undergraduate psychology course that students in teacher-assigned heterogeneous cooperative learning groups scored higher on final course examinations than students in self-selected groups. However, this difference did not reach statistical significance, possibly due to the limited size of the sample (there were only eighty-five students in the study) or the fact that, unlike the Lawrenz and Munch study, homogeneous groups were not part of the design.

That brings us to the key concern of the present study: differences between homogeneously and heterogeneously grouped cooperative learning groups. This is one area in which there has been little research even at the elementary school level, perhaps because of the strong assumption in cooperative learning about the importance of heterogeneous groups to promote social interaction among groups that often do not interact under less-structured arrangements. A primary concern has been to "maximize heterogeneity of skills and abilities," as well as to mix together as much as possible different social groups, with the goal of "promoting more favorable evaluations of outgroup members" (Miller and Harrington 1992, 212). The use of homogeneous cooperative learning groups has therefore been rather unusual. In fact, almost all of the cooperative learning methods promoted in the past three decades—and almost all of the research on the effectiveness of cooperative learning—have used heterogeneous grouping (Slavin 1981; Watson and Marshall 1995; Woolfolk 1998).

For this reason, even though almost all of its adherents have consistently emphasized the importance of heterogeneous grouping in cooperative learning, there is actually little evidence of its effectiveness in comparison to homogeneous grouping. What little research that has been done at the elementary and secondary levels suggests a pattern similar to that found in noncooperative learning settings: high achievers do much better in homogeneous groups; among average and low achievers there is little difference between students in heterogeneous and homogeneous groups; and high-achieving students frequently have a poor attitude toward group work (Allan 1991; Engelhard and Monsaas 1989; Kulik and Kulik 1982, 1984; Loveless 1999; Slavin 1990a; Sternberg and Willard 2002).

What about the effects of heterogeneous versus homogeneous grouping in undergraduate courses employing cooperative learning? Two studies, both completed in college science classrooms, made such comparisons, and although both (Lawrenz and Munch 1984; Watson and Marshall 1995) found that the homogeneously grouped students achieved at somewhat higher levels than the heterogeneously grouped students, neither comparison reached the .05 level of statistical significance. O'Donnell and Dansereau (1992) presented evidence suggesting that when their highly structured "scripted cooperation" model, in which dyads read and review textual material together, is used, lower-ability college students achieve more in heterogeneous groups than in homogeneous groups.

For the present study we compared the impact of heterogeneous and homogeneous cooperative learning groups on achievement in an undergraduate educational psychology course. The sample size was larger than those used in the studies by both Watson and Marshall (1995) and Lawrenz and Munch (1984), and the cooperative learning model employed was more typical of the kind of cooperative learning used in college classrooms than the highly constrained "scripted cooperation" model of O'Donnell and Dansereau (1992). The purpose was to compare the effects on achievement of placement in homogeneously versus heterogeneously grouped cooperative learning groups. The study was designed to allow overall comparisons of achievement as well as comparisons of the achievement of low-, average-, and high-achieving students in the two kinds of groups.

Method

The study was conducted over the course of three semesters using 137 students in six sections of an educational psychology class at a private liberal arts college with a large teacher-preparation program. The students in the study had average combined SAT scores of 1060, with a minimum GPA of 2.50 (which was required for admission to the course). The sample was largely white (94 percent) and female (91 percent). Ninety-three percent of the students were sophomores (the rest were juniors), and although their exact ages were not recorded, more than 90 percent were traditional students and therefore about nineteen or twenty years old.

Of the students who participated, sixty-eight were assigned to homogeneously grouped teams and sixty-nine were assigned to heterogeneously grouped teams based solely on early achievement in the course. In each section of the class, six teams were formed of three or four students each based on the results of a multiple-choice quiz given during the fourth week of the fourteen-week course. This was one of three graded tests given during the semester that together determined 60 percent of students' grades. (The remaining 40 percent of the grade was based on papers, the grades of which were not used in this study. Students in each pair of classes had identical paper assignments, however.) Prior to this quiz, students had been randomly assigned to a team, and one week after the quiz (with no connection suggested between the new team groupings and the quiz) new teams were formed. All students, therefore, worked for the first five weeks of the course in randomly assigned cooperative learning groups and for the final nine weeks in cooperative learning groups that were either homogeneous or heterogeneous in terms of achievement (based on scores on the first quiz).

Grouping was achieved as follows: In each class, the eight students who scored highest on the quiz were identified (for the purpose of this study only) as high-achieving students, and the eight who scored lowest were identified as low-achieving students. The remaining students were identified as average achievers. This identification was used to place students in groups in the homogeneously grouped classes, each of which had two high-achieving, two average-achieving, and two low-achieving teams. In the heterogeneously grouped sections, students were put into groups using the following formula: one high achiever, one low achiever, one average achiever, and (if the team had four members) one more student who could be from any group. These groups remained in place until the end of the course. Four students dropped the course after the teams had been formed: two low-achievers in heterogeneous groups and two average achievers, one from a heterogeneous group and the other from a homogeneous group.

Each semester for three semesters, the same professor taught two sections of an educational psychology course, which is open to all students but required of education majors. Each section was self-contained and met for ninety minutes twice a week for fourteen weeks. Class size was capped at twenty-five students. The syllabus used each semester was the same for both sections, as were all assignments, lectures, and activities, and students were assigned to sections on an essentially random basis. (Students signed up for one section or the other, but the meeting times were similar—either 11 AM or 1 PM every Tuesday and Thursday—and students knew they would have the same professor in either class.) In both sections cooperative learning groups were employed regularly. A random selection procedure determined which section would use which kind of grouping each semester. The students were not aware that this study was being conducted, nor were they aware how the cooperative learning groups were formed. They were to be told, if they asked, that they were grouped randomly. No one asked, however, in any of the sections.

Although the groups were sometimes referred to both as "teams" and "groups" in class, there was no competition among them for rewards of any kind, which previous research has shown are not needed when using cooperative learning in college classrooms (Baer and Baer 1996; Watson and Marshall 1995). The tasks carried out by the groups were all ungraded assignments that were completed during class time.

All teams were given the same assignments. A typical assignment was to discuss a question raised by the professor. These questions generally asked students to do such things as (*a*) simply think and share ideas about some topic (e.g., "How are Piaget's concept of equilibration and Vygotsky's idea of the zone of proximal development alike, and how are the two concepts different?"); (*b*) analyze an argument, critique a research design, or evaluate a conclusion (e.g., after reading or hearing about a particular study, students were sometimes asked to look for

design flaws or problems with the authors' conclusions based on possibly insufficient evidence); (*c*) apply a concept (e.g., after studying different kinds of instructional objectives, students were asked to develop a set of cognitive and behavioral objectives for a particular topic); or (*d*) simply review a topic by asking one another questions about it. Sometimes teams were asked to develop an answer to be shared with the rest of the class, and sometimes the activity ended after the groups had discussed the topic among themselves (with no reporting to either the class or the professor).

Although teams were occasionally asked to turn in a written response of some kind, this was unusual, and the teams' products were never graded. All group activities were conducted during class and for the same length of time. Because students were placed in classes randomly and random assignment was used to determine which classes received homogeneous or heterogeneous grouping, extraneous variables (e.g., industriousness, previous knowledge of and interest in the subject matter, personality traits, study habits, learning styles, etc.) can be safely assumed to be equal between the two groups.

Results

As can be seen in table 1, there was no significant difference between the two groups' scores on the multiple-choice quiz, which were the basis for placement in teams and prior to which all students were in randomly assigned cooperative learning groups. The two groups thus appeared very similar prior to the beginning of the study, which started when students were placed in either heterogeneous or homogeneous groups.

The outcome measures were the multiple-choice portions of the midterm and final examinations. The homogeneous group outperformed the heterogeneous group significantly ($p < .04$) on the final examination (at the end of the fourteen-week course and after nine weeks of working with either a homogeneously or heterogeneously grouped team) and at a level of borderline statistical significance ($p < .10$) on the midterm examination (nine weeks into the course and after just four weeks of working with either a homogeneously or heterogeneously grouped team).

TABLE 1. Mean Scores—All Students

Test	Homogeneous Group ($N = 68$)	Heterogeneous Group ($N = 69$)	F Ratio (1, 136)	Prob > F
Quiz	74.1 (s.d. = 14.56)	75.0 (s.d. = 15.84)	0.062	.804
Midterm	85.6 (s.d. = 14.01)	81.9 (s.d. = 11.70)	2.825	.095
Final	84.4 (s.d. = 14.60)	79.3 (s.d. = 13.66)	4.439	.037

Note: The F ratio is a way of assessing the likelihood that a difference between group means is statistically significant. The larger the F ratio, the greater the statistical significance. The likelihood that an observed difference may be due to chance is expressed as a percentage in the Prob > F column. Any value in the Prob > F column less than .05 is generally considered statistically significant (that is, such differences are assumed to represent real differences that are not due merely to chance effects).

TABLE 2. Mean Scores—Low Achievers

Test	Homogeneous Group ($N = 24$)	Heterogeneous Group ($N = 22$)	F Ratio (1, 45)	Prob > F
Quiz	54.6 (s.d. = 6.77)	52.5 (s.d. = 10.31)	0.427	.517
Midterm	74.3 (s.d. = 13.19)	73.5 (s.d. = 12.33)	0.044	.834
Final	69.5 (s.d. = 10.36)	69.6 (s.d. = 11.44)	0.002	.966

TABLE 3. Mean Scores—Average Achievers

Test	Homogeneous Group ($N = 21$)	Heterogeneous Group ($N = 22$)	F Ratio (1, 42)	Prob > F
Quiz	74.9 (s.d. = 3.36)	76.5 (s.d. = 2.69)	1.892	.176
Midterm	87.9 (s.d. = 12.51)	82.2 (s.d. = 7.21)	3.249	.079
Final	93.3 (s.d. = 9.89)	84.7 (s.d. = 9.00)	8.929	.005

TABLE 4. Mean Scores—High Achievers

Test	Homogeneous Group ($N = 24$)	Heterogeneous Group ($N = 24$)	F Ratio (1, 42)	Prob > F
Quiz	93.3 (s.d. = 7.88)	94.4 (s.d. = 8.69)	0.121	.730
Midterm	94.8 (s.d. = 9.73)	89.1 (s.d. = 5.71)	6.185	.017
Final	91.8 (s.d. = 8.38)	82.9 (s.d. = 14.83)	7.109	.011

Tables 2, 3, and 4 list scores on the various tests for low-, average-, and high-achieving students. Again, the outcome measures were the final and midterm examination grades; quiz grades have been provided only to show that there were no significant differences in the two groups prior to the onset of the experiment. For low achievers there was essentially no difference between those working in homogeneous and heterogeneous groups on either the midterm or final examinations. For both average and high achievers, however, homogeneous grouping was clearly preferable, leading to substantial (and statistically significant) differences on the

final examination among both groups. Even on the midterm examination, which was given after students had worked with their assigned groups for only four weeks, homogeneously grouped high achievers significantly outperformed heterogeneously grouped high achievers, and among average achievers, the homogeneous group outperformed the heterogeneous group at a level approaching statistical significance ($p < .08$).

Discussion

These results suggest that when using cooperative learning, homogeneous grouping in an undergraduate course results in higher achievement than heterogeneous grouping, at least in a course structured in a way that performance on group work is not graded and group work is conducted during class time. The levels of statistical significance obtained strongly suggest that these differences were not the result of chance, but rather represent real differences. Because the only systematic difference between the groups was the kind of grouping (homogeneous versus heterogeneous) used, it is reasonable to conclude with some confidence that the differences in how students were grouped caused the observed differences in achievement. It is, of course, possible that the results might differ if one used different cooperative learning procedures (e.g., if group assignments were carried out during nonclass time or if the work of groups was graded in some way).[1]

The measures employed to assess student achievement in the class were not limited to the multiple-choice exams used in this study for comparison purposes. But use of only the objectively scored tests avoided the possibility that any expectations that the professor may have had could have influenced the outcome (as might happen were grades on papers or essay test questions used to compare group achievement). As in other research on grouping methods (whether cooperative learning is employed or not), the study was not conducted in a manner that ensured that the professor would not know the nature of the groups (Loveless 1999), and it is therefore impossible to be certain that the professor's expectations in no way influenced the outcome. Random assignment, use of only objectively scored achievement measures, and identical assignments and in-class activities, however, represent significant control over extraneous variables that might have influenced the outcome, and together they provide about as much control as is generally possible in ecologically valid educational research.

The results are significant because they run counter to the common (and often recommended) practice, among college professors who employ cooperative learning, of grouping students heterogeneously (e.g., Halpern 2000). In college classrooms with fairly wide ranges of student ability, homogeneous grouping could result in significant achievement gains, at least among average- and high-achieving students, while doing no harm to the achievement of low-achieving students.

This study does not tell why homogeneously grouped students outperformed heterogeneously grouped students. One possibility that has been suggested for similar outcomes in research with younger students is that conversations among students with similar levels of knowledge and understanding of the topic may be more interesting to the students and more likely to take place at a level appropriate to the knowledge and skills of the students involved (Loveless 1999; Webb 1992). If this interpretation is correct, then it is likely that these results would be generalizable to other college courses in which there is a wide range of knowledge of the subject being studied and plentiful opportunity for in-class cooperative learning activities.

There may be other advantages to heterogeneous grouping of students, of course, such as the promotion of intergroup relations. Baer and Baer (1997) demonstrated that assigning college students to cooperative learning groups to promote heterogeneity in terms of the racial, ethnic, and gender make-up of groups may have a positive impact on student achievement in college cooperative learning groups, in comparison to self-selected groups. Lawrenz and Munch (1992) demonstrated a clear superiority of teacher-assigned groups to self-selected groups when cooperative learning was applied in a college science laboratory. There is no reason why cooperative learning groups cannot be homogeneous with respect to academic achievement and heterogeneous with respect to other student characteristics, including gender, race, or ethnicity. Such a practice would allow professors to promote important nonacademic goals of cooperative learning, such as improving intergroup relations, while grouping students in ways most likely to result in the highest levels of student achievement.

Key words: cooperative learning, homogeneous and heterogeneous grouping, achievement levels

NOTE

Grading group projects of any kind and including such grades as part of individual student grades raises questions about fairness that are beyond the scope of this paper, but it should be noted that using cooperative learning group grades in such a manner (or rewarding team performances in any way) would raise especially thorny ethical issues were the groups structured using homogeneous grouping, as would using group rewards with self-selected groups. One cannot easily brush aside concerns about the fairness of basing individual grades, even in part, on the work of other students regardless of the method of grouping, but homogeneous grouping makes this fundamental issue of fairness especially salient.

REFERENCES

Allan, S. D. 1991. Ability-grouping research reviews: What do they say about grouping and the gifted? *Educational Leadership* 48 (6): 11–16.

Baer, J., and S. K. Baer. 1996. Are group rewards necessary for cooperative learning? In *Selected papers from the Seventh National Conference on College Teaching and Learning*, 9–14. Jacksonville: Florida Community College at Jacksonville.

———. 1997. Group structure in cooperative learning. Paper presented at the Eighth National Conference on College Teaching and Learning, Jacksonville, Fla.

Engelhard, G., and J. Monsaas. 1989. Academic performance, gender, and the cooperative attitudes of third, fifth, and seventh graders. *Journal of Research and Development in Education* 22 (2): 13–17.

Halpern, D. F. 2000. Creating cooperative learning environments. *American Psychological Society Observer* 13 (3): 14–31.

Hertz-Lazarowitz, R., and N. Miller, eds. 1992. *Interaction in cooperative groups: The theoretical anatomy of group learning*. Cambridge: Cambridge University Press.

Johnson, D., and R. Johnson. 1989. *Cooperation and competition: Theory and research*. Edina, Minn.: Interaction.

Kulik, C. L., and J. A. Kulik. 1982. Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal* 19:415–28.

———. 1984. Effects of ability grouping on elementary school pupils: A meta-analysis. Paper presented at the annual meeting of the American Psychological Association, Toronto. ERIC No. ED 255 329.

Lawrenz, F., and T. W. Munch. 1984. The effect of grouping of laboratory students on selected educational outcomes. *Journal of Research in Science Teaching* 21:699–708.

Loveless, T. 1999. *The tracking wars*. Washington, D.C.: Brookings Institution Press.

O'Donnell, A. M., and D. F. Dansereau. 1992. Scripted dyadic cooperation. In *Interaction in cooperative groups: The theoretical anatomy of group learning*, edited by R. Hertz-Lazarowitz and N. Miller, 120–41. Cambridge: Cambridge University Press.

Slavin, R. E. 1980. Cooperative learning. *Review of Educational Research* 50:315–42.

———. 1981. Synthesis of research on cooperative learning. *Educational Leadership* 38:655–60.

———. 1984. Students motivating students to excel: Cooperative incentives, cooperative tasks, and student achievement. *Elementary School Journal* 85 (1): 53–63.

———. 1990a. Ability grouping and student achievement in elementary schools: A best-evidence synthesis. Report No. 1, Center for Research on Elementary and Middle Schools, Johns Hopkins University, Baltimore.

———. 1990b. *Cooperative learning: Theory, research, and practice*. Englewood, N.J.: Prentice-Hall.

———. 1991. Group rewards make group-work work. *Educational Leadership* 48 (5): 89–91.

———. 1992. When and why does cooperative learning increase achievement? In *Interaction in cooperative groups: The theoretical anatomy of group learning*, edited by R. Hertz-Lazarowitz and N. Miller, 145–73. Cambridge: Cambridge University Press.

Sternberg, R. J., and W. M. Willard. 2002. *Educational psychology*. Boston: Allyn and Bacon.

Tobias, S. 1989. Tracked to fail. *Psychology Today* September, 54–60.

Watson, S. B. 1991. Cooperative learning and group educational modules: Effects on cognitive achievement of high school biology students. *Journal of Research in Science Teaching* 27:671–83.

Watson, S. B., and J. E. Marshall. Effects of cooperative incentives and heterogeneous arrangement on achievement and interaction of cooperative learning groups in a college life science course. *Journal of Research in Science Teaching* 32:291–99.

Webb, N. M. 1992. Testing a theoretical model of student interaction and learning in small groups. In *Interaction in cooperative groups: The theoretical anatomy of group learning*, ed. R. Hertz-Lazarowitz and N. Miller, 102-19. Cambridge: Cambridge University Press.

Woolfolk, A. 1998. *Educational psychology*. 7th ed. Boston: Allyn and Bacon.