Review of Educational Research
Winter 1996, Vol. 66, No. 4, pp. 423–458

Within-Class Grouping: A Meta-Analysis

Yiping Lou
Philip C. Abrami
Concordia University
John C. Spence
University of Alberta
Catherine Poulsen
Bette Chambers
Sylvia d'Apollonia
Concordia University

The effects of within-class grouping on student achievement and other outcomes were quantitatively integrated using two sets of study findings. The first set included 145 effect sizes and explored the effects of grouping versus no grouping on several outcomes. Overall, the average achievement effect size was +0.17, favoring small-group learning. The second set included 20 effect sizes which directly compared the achievement effects of homogeneous versus heterogeneous ability grouping. Overall, the results favored homogeneous grouping; the average effect size was +0.12. The variability in both sets of study findings was heterogeneous, and the effects were explored further. To be maximally effective, within-class grouping practices require the adaptation of instruction methods and materials for small-group learning.

Contemporary classrooms are notable for the number and diversity of students who occupy them. Economic pressures in many regions have resulted in increased class sizes. Detracking or destreaming, the mainstreaming of students with special needs, and the reduction of special programs for gifted students make it likely that teachers face students who have a broad spectrum of needs, abilities, goals, and interests and who may differ along racial, ethnic, linguistic, and economic lines. The mosaic of students who populate classrooms means that teachers face difficult pedagogical decisions if students are to learn effectively and enjoyably. One decision concerns whether to group students for instruction within class and teach them accordingly.

The term *small-group instruction* has different meanings. In the loosest sense it means the physical placement of students into groups for the purposes of

The authors express their appreciation to Professor Anne Wade, who conducted the electronic searches used to identify the studies included in the review. This research was supported by grants to the authors from the Social Sciences and Humanities Research Council (Government of Canada) and the Fonds pour la formation de chercheurs et l'aide a la recherche (Government of Quebec). Portions of this article were presented at the Annual Meetings of the American Educational Research Association in San Francisco (April, 1995) and New York (April, 1996).

learning. Gamoran (1987) refers to this placement as the organizational structure of the classroom. In the strictest sense, however, small-group instruction means the use of specific instructional strategies when students are placed together to learn. Gamoran refers to this as the instructional processes occurring within classes and groups. In this article we will use the term *small-group instruction* in the loosest sense and interchangeably with *within-class grouping* and *intraclass grouping,* although we also attempt to explore the influence of instructional strategies on small groups.

The effects of within-class grouping have been the focus of educational research for some time. However, individual study findings appear quite varied and therefore call for a careful and systematic review. Previous meta-analyses (Kulik & Kulik, 1987, 1991; Slavin, 1987)[1] explored whether within-class grouping was superior to whole-class instruction in promoting student learning. While positive effects of grouping were reported in each review, the average effects were not consistent in size. Slavin (1987) reported an average effect size of +0.32; Kulik and Kulik (1987) reported an average effect size of +0.17, while Kulik and Kulik (1991) reported an average effect size of +0.25. Furthermore, the samples of studies included were small and varied from review to review (7 in Slavin, 1987; 15 in Kulik & Kulik, 1987; 11 in Kulik & Kulik, 1991). In addition, the reviews concentrated on the effects of within-class grouping versus no grouping; the effects of homogeneous grouping and heterogeneous grouping were not compared. Finally, no review determined whether the set of findings was uniform or probed the effects to determine when grouping was best and under what conditions.

Therefore, the primary objective of this study is to integrate quantitatively the research findings on the effects of within-class grouping and to explain study-to-study variability in outcomes. Specifically, we ask three broad questions about grouping. First, how much, if any, does placing students in small groups facilitate their learning and other outcomes, such as attitudes and self-concept? Second, which factors explain variability in study findings? Third, which type of grouping is best and under what conditions? In sum, our meta-analysis is designed to be more comprehensive and thorough than previous reviews.

Arguments for Whole-Class Instruction

Whole-class instruction means that students are taught as a single, large group. In whole-class instruction, there is an emphasis on the uniformity of instruction rather than on the diversity of instruction. For example, it is common for the teacher utilizing whole-class instruction to provide a single, detailed explanation to the class followed by the assignment of individual seatwork. In whole-class instruction, the emphasis is on teacher explanations and encouragement, rather than on peer explanations and encouragement, to promote student learning. There are several reasons for utilizing whole-class instruction. First, the uniformity of instruction means that teachers may spend preparation time on developing a single set of instructional materials appropriate to the content to be learned rather than spend time developing many sets of materials. Second, whole-class instruction means that teachers may emphasize a single set of instructional objectives for all students, objectives which are sometimes encountered in a required or core curriculum. The adherence to specified objectives also implies a fixed pace of instruction for all students. Third, teachers may utilize their content and pedagogical expertise to explain new material orally to all students. Students may use seatwork to practice the skills taught by the instructor and to explore content further. Thus, learning may be facilitated by direct instruction from the teacher (since instructional time is maximized) and by guided, individual practice. Fourth, students may be motivated to learn by tangible or symbolic incentives provided by the teacher, which sometimes places students in competition with one another to excel. Fifth, whole-class instruction means that all students may be exposed to the same learning opportunities, emphasizing the open, democratic principles of the educational system and the realities of life in a world that operates according to the survival of the fittest.

Arguments for Small-Group Instruction

Small-group instruction means that a class of students is taught in several small groups. In small-group instruction, there is often an emphasis on diversity of instruction rather than on uniformity of instruction. The teacher may provide either a single, brief explanation to the class as a whole or give different instructions to each group. The teacher may either assign the same seatwork to each group or vary the assignment from group to group. In small-group instruction, peer helping is often encouraged to promote student learning. There are several reasons for utilizing small-group instruction. First, the emphasis on peer learning means that the teacher may have more time to provide either remedial assistance to students experiencing difficulties or enrichment activities to students who have already mastered prescribed content. Second, using within-class grouping means that teachers may have greater flexibility in adjusting the learning objectives and the pace of instruction to meet individual learning needs. Using homogeneous ability groups means that the teacher can increase the pace and level of instruction for high achievers and provide more individual attention, repetition, and review for low achievers. Third, students in small groups may engage in such activities as orally rehearsing material, explaining material to others, discovering solutions, and debating and discussing content and procedural issues. Thus, teachers may capitalize on the social aspects of cognitive growth (Piaget, 1954; Vygotsky, 1978) emphasizing the development of higher-order thinking skills. Fourth, students who learn together in small groups may be motivated by cooperative, as opposed to competitive, incentive structures. Fifth, small-group instruction means that students may have the opportunity to develop social and communication skills because of the need and opportunity to work with others to learn.

Within-Class Grouping Then and Now

While classroom teachers have been dividing students into learning groups for some time, the research evidence is more recent. Few studies existed at the time of Petty's (1953b) review, but the situation changed thereafter. In the late 1950s and 1960s, researchers most often examined the effects of homogeneous within-class grouping where teachers would divide a class into subgroups for specific activities and purposes, especially for elementary instruction in reading and sometimes for instruction in mathematics.

In a typical study, Dewar (1963) explored the effects of within-class grouping on the learning of arithmetic among classes of sixth-grade students. Each class was divided into three groups on the basis of test results, school records, and teacher judgment. Membership in the groups remained constant for the term, and teachers used specially prepared materials along with textbooks from Grades 4–8. The highest-scoring students used sixth-, seventh-, and eighth-grade materials; the middle-scoring groups used fifth-, sixth-, and seventh-grade materials; and the lowest-scoring groups used fourth-, fifth-, and sixth-grade materials. Each teacher spent almost an hour per day on mathematics instruction. The teacher presented material to a group for approximately 15 minutes before moving on to another group. While the teacher was presenting material to one group, all other groups worked on their assignments.

By the late 1970s and 1980s, researchers were again examining the effects of within-class grouping, but with a particular emphasis on cooperative learning. In many of these studies, teachers would most often divide students into heterogeneous ability groups.

Cooperative learning is one form of small-group instruction which utilizes both positive interdependence and individual accountability to encourage students to learn (Abrami et al., 1995). Positive interdependence exists when individual accomplishments contribute positively to the accomplishments of others—for example, when the members of a group all receive the same recognition for the group's accomplishments. Individual accountability exists when students are responsible for their own learning and the learning of other group members—for example, when each member of a group has clear tasks or roles to accomplish.

In a typical study, Bejarano (1987) compared the effects of a cooperative learning technique known as Student Teams and Achievement Divisions (STAD) with those of whole-class instruction. The STAD technique had several components: (a) assignment of students to teams, where each team consisted of members who were heterogeneous in ability; (b) brief, whole-class instruction; (c) team study using worksheets common to all teams; (d) individual quizzes; (e) calculation of quiz results in the form of improvement points noting personal gains; and (f) team recognition based on the combination of individual improvement points. Of particular note, the STAD technique emphasized both an individual's responsibility for his or her own learning and the importance of contributing to the learning success of teammates.

Duration and Composition of Within-Class Groups

Teachers may use small groups for only a portion of class time, for only a fraction of a semester, or for selected activities. For example, Coldiron, Braddock, and McPartland (1987) found that almost all (about 90%) elementary school students in Pennsylvania were exposed to within-class ability grouping for reading, but only about one third of the students were exposed to grouping in mathematics. These differences in pervasiveness and duration are worth exploring to find the optimal conditions for small-group instruction and to test for the presence of novelty effects.

Grouping for instruction can be accomplished in a variety of ways. Students may be self-assigned, randomly assigned, or instructor assigned to groups. Groups may be formed on the basis of (a) common interests, common skills, or friendships or (b) diverse interests, diverse skills, or unfamiliarity. Groups formed on the basis of common interests may be intrinsically motivated to work together to learn.

Homogeneous ability groups may encourage learning as students attempt to maintain a pace commensurate with other group members. Friendship groups may be motivated to learn by the effects of group cohesiveness. In contrast, groups formed on the basis of diverse interests may enhance learning if multiple perspectives facilitate goal attainment or stimulate creative controversy. Heterogeneous ability groups may foster learning through the use of elaborated explanations whereby the more able students tutor the less able ones. Placing strangers together may better integrate all students into the classroom milieu, avoid outcasts, and teach students tolerance, acceptance, and strategies for working successfully with a diversity of partners.

Clearly, one issue underlying group composition is whether or not groups should be heterogeneously composed according to ability, interests, liking, gender, ethnicity, and so on. Unfortunately, a scant number of studies exist to integrate findings on group composition criteria other than relative ability or prior achievement. This may be satisfactory if relative ability is the most salient quality to consider when forming groups. Wilkinson (1986), however, concluded that teachers need to consider many factors when they assign students to instructional groups and that individual students' needs and characteristics, such as developmental level, are the most important.

In the absence of sufficient primary research, this integration is restricted to an examination of the extent to which small groups composed of students who are relatively similar in ability or prior achievement achieve in comparison to small groups composed of students who are relatively dissimilar in ability or prior achievement. In addition, the integration explores the extent to which type of ability grouping affects learning for students relatively high in ability, medium or average in ability, and low in ability.

A second issue in group composition concerns the nature of the task required of students learning in small groups. Noddings (1989) argued that the composition of classroom groups interacts with the nature of the learning task to affect student learning. For many academic tasks, Noddings noted that teachers frequently assign students to homogeneous ability groups, such as traditional reading groups.

In traditional reading and other group activities, student placement can be relatively stable and long-term, depending almost entirely on the results of initial achievement or ability testing (Hallinan, 1984). Across homogeneous groups, the nature of the task and pace of learning often require adjustment to suit the learners. If adjustments are not made it is possible that (a) some groups will be unable to accomplish the task because no member has the requisite skills for learning, and (b) some groups will too readily accomplish the task because all members have already acquired the skills which the task was designed to develop. Indeed, Kulik and Kulik (1991) concluded that homogeneous within-class grouping would be pointless without adapting the curriculum materials to the needs and abilities of students in each group.

Noddings (1989) also noted that when the task is a typical academic one and groups are heterogeneously formed, the group members often turn to the most able student for help. When this happens, there may be little interaction and limited engagement, thereby minimizing understanding even though students appear to get the answer right. Webb (1989) also found that merely giving or receiving answers was insufficient and could interfere with learning if and when such behavior circumvented cognitive engagement. Giving and receiving elaborated explanations, on the other hand, was positively related to achievement.

Methodological Considerations

Good and Marshall (1984) commented on the tendency for studies to manipulate factors along with small-group instruction. For example, curriculum and teaching method were not always held constant across conditions. In some studies, teachers were requested to keep content and teaching method the same. In other studies, special materials and pace were made available for students according to their group placement. In yet other studies, teachers using small groups were given special training, while control teachers using whole-class instruction were not.

One solution to the problem of confounded effects is to exclude studies where grouping method varies with other factors. Slavin (1987) dealt with the confounding problem in exactly this way. His review of studies of gifted and special education students excluded studies in which curriculum, class size, resources, and goals varied along with the between-class grouping plan. He found no large effects in the studies that remained. Critics (e.g., Allan, 1991) responded that this finding was used by educational administrators to make decisions about all special education and gifted programs, including those not studied. For the comparison of small-group instruction to no grouping, we did not exclude studies in which grouping practices, instruction, and curriculum varied. Instead, we attempted to identify, code, and analyze the effects of factors that varied along with the use of small-group instruction.

Policy Issues in Grouping Students for Instruction

There are strong and emotional arguments both for (e.g., Allan, 1991) and against (e.g., Oakes, 1985) between-class grouping. Many of the arguments about group composition between classes apply to within-class grouping. For example, there is a concern that it is unethical to stigmatize low-ability students by using a system of within-class tracking whereby students are segregated according to ability or prior achievement. In particular, there is the fear that low-ability students placed in homogeneous ability groups will be denied opportunities to learn and be unmotivated to learn because of peer, personal, and teacher expectations of poor performance.

On the other hand, there is a concern that it is unethical to retard the achievement of high-ability students by assigning them to heterogeneous within-class groups, in which students are integrated according to ability. In particular, there is the fear that high-ability students placed in heterogeneous groups will be denied opportunities to learn because much of the material has already been mastered by them, because the pace of learning in the group is below their capacity, and because their role in the group is not to learn but to instruct less able students.

No review of empirical research can resolve differences in educational policy which arise from differences in educational values and philosophy. For example, in an early narrative review of the literature on within-class grouping, Petty (1953b) argued that grouping students for instruction was then viewed as "a democratic instructional procedure designed to adapt the curriculum and learning environment to the abilities and needs of individual pupils and to provide appropriate means for fostering their continuous development" (p. 7). More recently, others (e.g., Rosenbaum, 1976) have argued that the use of ability grouping may serve to increase divisions along ethnic, racial, and class lines.

Still others (Good & Marshall, 1984; Kulik & Kulik, 1987) have noted that these values change over time. For example, during the 1950s "excellence" was a byword in education, and between-class grouping was seen as beneficial for high-ability students. In the 1960s and 1970s, concerns about equal opportunity increased, and between-class grouping was seen as harmful for disadvantaged students. What a review can do is help inform policy by exploring the empirical basis of beliefs which underlie a particular philosophy and by suggesting directions for future research if the evidence is lacking. The present review was undertaken, in part, with this purpose in mind.

Method

This meta-analysis primarily examined the effects of within-class grouping on student achievement at the elementary, secondary, and postsecondary levels. However, outcomes other than achievement were included, when they were available, and analyzed separately. Two groups of studies were identified and analyzed independently. Analysis 1 included studies which compared within-class grouping with no grouping. Analysis 2 included studies which directly compared homogeneous within-class grouping with heterogeneous within-class grouping. The procedures employed to conduct the quantitative integrations are outlined below under the following headings: Literature Search, Inclusion/Exclusion Criteria, Study Features Coding, Number of Findings Extracted, Effect Size Calculations, and Data Analyses.

Literature Search

The studies used in this meta-analysis were located via a comprehensive search of the literature. Electronic searches were performed on the ERIC (1966–1994), PsycLIT (1974–1994), Sociofile (1974–1994), Dissertation Abstracts (1965–1994), and Social Sciences Citation Index (1989–1994) databases. Although the search strategy varied depending on the database, search terms included *group composition, grouping for instructional purposes, small group, team learning, team instruction, heterogeneous/homogeneous grouping, group structure, ability grouping, peer tutoring,* and *cooperative learning.* Through branching from primary studies and review articles, other citations were found and included. In total, the search uncovered over 3,000 published articles concerning within-class grouping.

Inclusion/Exclusion Criteria

To be included in this review a study had to meet the following criteria.

(1) The research must have occurred within the classroom at the elementary, secondary, or postsecondary school level. Consequently, within-class grouping studies of organizational behavior (e.g., in a business setting) or social psychology (e.g., group therapy) were excluded.

(2) The research had to involve within-class ability grouping, either homogeneous ability grouping or heterogeneous ability grouping. The same teacher had to instruct a classroom of groups that were working in close proximity

to one another. Studies were not included if they involved between-class tracking or grouping in a laboratory setting outside the classroom.

(3) The minimum group size was 2 students, and the maximum group size was 10 students.

(4) Grouping had to be in place for more than 1 day. Thus, studies involving only one session (e.g., a 1-hour lesson) were not included.

(5) If training of any kind was offered to the students, then all group members must have received it.

(6) The research had to report measured outcomes from both experimental and control groups. Studies reporting only single-group pretest and posttest comparisons were initially coded, but the results were eventually excluded for two reasons. First, the methodological shortcomings inherent in this design make unambiguous interpretations of outcomes tenuous. Second, the gain score results from these studies were significantly different from the results of studies employing experimental and control groups. One-shot case studies were also excluded a priori on methodological grounds.

(7) Achievement, attitudes, and self-concept were included, but each was analyzed separately. Classroom behaviors and interrelationships among measures were initially included but not analyzed for this meta-analysis.

(8) Finally, research primarily involving either children with learning disabilities or enrichment programs for gifted students were excluded.

The searches identified more than 500 studies on cooperative learning (cf. Johnson & Johnson, 1989), most employing heterogeneous grouping. In order to balance the number of grouping studies using this instructional method with grouping studies using other instructional methods, we included only those studies identified by the search term *cooperative learning* and the term *homogeneous/heterogenous grouping* or *ability grouping* or *group composition.*

Most studies on peer tutoring were excluded because we considered the group dynamics of tutoring to be quite different from other grouping methods. Members of peer tutoring dyads often functioned as learner (tutee) and teacher (tutor), effectively creating learning groups of one. Also, tutors often received special training and extra rewards (e.g., free time during a regular class, extra credit, etc.). Therefore, studies of cross-age/cross-class tutoring, tutoring outside classrooms or for remuneration, and one-way tutoring (i.e., one tutor taught one student) were excluded. However, paired learning and reciprocal tutoring, in which two students alternated between the roles of tutor and tutee, were included.

Studies on group-based mastery learning were included if they met the inclusion/exclusion criteria (e.g., students were placed together in small groups for the purposes of learning, group size smaller than 10, etc.). Thus, group-based mastery learning studies were excluded if students did remediation or enrichment work individually.

Using the above criteria, abstracts from electronic searches and references from primary studies and reviews were first examined by two researchers to identify potential studies to include. Differences between the independent judgments of the two researchers were resolved through discussion. If there was a doubt, the study was collected. Next, the collected studies were each read by two researchers for possible inclusion. Any study that was considered for exclusion by one researcher was cross-checked by the other researcher. Sixty-six studies met the inclusion criteria and were included in this review. Table 1 lists the studies included in this meta-analysis, those included in previous meta-analyses, and the reasons for study exclusions. Six studies that were included in Slavin (1987, 1990) or Kulik and Kulik (1987, 1991) were excluded for failing to meet one or more of the criteria specified here.

Study Features Coding

Literature reviews are often unnecessary when the findings on a phenomenon appear uniform across studies. Instead, it is the apparent inconsistency of findings that motivates the search for factors which explain differences among studies. The purpose of coding study features is to identify those methodological and substantive characteristics which may be responsible for significant variations in the findings. In this review, nomological coding was used first to objectively identify salient study features in the literature, thereby avoiding reviewer bias (Abrami, Cohen, & d'Apollonia, 1988; Abrami, d'Apollonia, & Cohen, 1990). Features with more than three substantive occurrences in a random sample of 25% of the primary literature and review articles were included for nomological coding. Forty-five study features were initially identified and were organized into three major categories: outcome features, methodological features, and substantive features. Outcome features were generally concerned with the nature of the dependent measures analyzed in the study. Methodological features were generally concerned with the quality of the research design and the fidelity of the treatment. Substantive features included grouping characteristics, instructor and instructional characteristics, student characteristics, setting, and scope. Unfortunately, many features had too few cases to be included in the analyses and were subsequently dropped. (See Table 2 for a description of the 26 study features that were coded and used in this review.)

Study features coding was performed independently by two coders. Their initial coding agreement was 88.24%. Disagreements between the two coders were resolved through discussion and further review of the disputed studies.

Number of Findings Extracted

Effect sizes were extracted and are reported separately for each major outcome category, primarily achievement but also student attitudes and self-concept. For each major outcome category, several effect sizes were often extracted from a single study as long as they were distinguishable at the level of study features (e.g., effect sizes reported separately for each of three grade levels). When not distinguishable by any study feature, effect sizes were averaged. Multiple effect sizes extracted from single studies can be problematic because methods of research integration normally assume that effect sizes are independent. This problem is, in our opinion, not especially pronounced when different subjects from a study provide separate measures of effect size. Such a condition arises in factorial designs, for example, when different effect sizes are extracted for male and female participants. In our analyses, each effect size was weighted by sample size, so that a study with two effect sizes based on 50 subjects each had the same overall weight as a study with a single effect size based on 100 subjects.

Dependence among effect sizes is pronounced when the same subjects from a study provide separate measures of effect size—a situation where within-group or repeated factors are present. For example, in this review we collected achievement data measured at mid-treatment, posttest, and delayed posttest. Ignoring the dependence and treating the within-group effects as independent would increase the Type I error rate of the homogeneity of effect size tests (i.e., tests of the variability in effect sizes) (Gleser & Olkin, 1994). Discarding the data would have the opposite effect, increasing the likelihood of committing a Type II error. In this meta-analysis, the problem was overcome by taking a single, random sample from the set of correlated effect sizes per feature for each affected study. This had the desirable effect of ensuring that all levels of a study feature were represented. For example, for the analysis of measure source, the selection of within-group findings was made randomly from among outcomes based on standardized tests, teacher-made tests, and researcher-made tests. This method was applied after all the study findings had been extracted and coded.

The study findings were extracted by two coders separately. The initial coding agreement on the number of findings to extract per study was 80.29%. Disagreements between the coders were resolved through subsequent discussion and further review of the disputed findings. Overall, there were 266 findings extracted prior to random sampling within studies. After random sampling, 165 independent findings were selected for analysis.

Effect Size Calculations

The basic index used for the effect size calculation was the mean of the experimental group minus the mean of the control group divided by the pooled standard deviation (PSD). The main reason for using the PSD is that often the assumption of homogeneity of variance in the population is reasonable, in which case the PSD is more stable and provides a better estimate of the population variance than the control group SD alone (Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Rosenthal, 1991). Another reason for the choice of the PSD is that estimated effect sizes based on incomplete results (e.g., t values, F values, ANOVA tables, or p levels) are more readily comparable to calculated effect sizes. Finally, the PSD is a more appropriate measure when it is unclear which group is the control condition, such as occurs when heterogeneous grouping and homogeneous grouping are compared.

In studies that report posttest data only, we used the posttest mean difference in the numerator and the posttest PSD in the denominator. In studies that provided gain scores or both pretest and posttest data, we used the gain score difference in the numerator to control for pretest differences, but the posttest PSD was used in the denominator rather than the gain score PSD since the gain score PSD is usually smaller than the posttest PSD (Glass, McGaw, & Smith, 1981). When the posttest SDs were not provided in the study, we tried to estimate the posttest PSD whenever possible. Such estimation requires r, which is, unfortunately, not usually reported in studies. In one case, we obtained $r = .88$ for the Stanford Achievement Test (1953 revision) from the Fifth Mental Measurements Yearbook (Gage, 1959). In other cases, we had to estimate a "typical" reliability for that class of measures based upon our knowledge of the literature. Specifically, we estimated $r = .85$ for standardized tests and $r = .75$ for unstandardized tests.

TABLE 1
Within-class grouping studies included in the previous and present meta-analyses

Studies	Slavin (1987)*	Slavin (1990)*	Kulik & Kulik (1987)*	Kulik & Kulik (1991)*	The present meta-analysis		
					Analysis	Outcome	Reason for excluding
Abu (1993)					1b	A, B	
Allen & VanSickle (1984)					1b	A, C	
Amaria et al. (1968)					1a, 1b, 2	A	
Armstrong (1993)					2	A, C	
Ballman (1988)					1b	A	
Bejarano (1987)					1b	A	
Berge (1990)					1b	A	
Bierden (1970)			x				II
Blaney et al. (1977)					1b	B, C	
Bright et al. (1980)					2	A	
Campbell (1964/1965)		x	x	x	1a	A	
Carter & Jones (1993)					2	A	
Chang (1993)					1b	A, B, C	
Cignetti (1974)			x	x	1a	A	
DeVries & Edwards (1973)					1b	B	
Dewar (1963)	x		x	x	1a	A	
Eddleman (1971)			x	x			III
Evans (1942)					1a	A	
Fantuzzo et al. (1990)					1a	A	
Hallinan & Sorensen (1985)					1a	A	
Harrah (1955/1956)		x	x		2	A	
Hay (1980/1981)					1b	A	
Heller & Fantuzzo (1993)					1b	A, B, C	
Hudgins (1960)					1b	A	

TABLE 1 (*continued*)

					The present meta-analysis		
Studies	Slavin (1987)*	Slavin (1990)*	Kulik & Kulik (1987)*	Kulik & Kulik (1991)*	Analysis	Outcome	Reason for excluding
Hulten & DeVries (1976)					1b	A, B, C	
Janicki & Peterson (1981)					1b	A, B	
Johnson et al. (1979)					1b	A, B	
Jones (1948)	x		x	x	1a	A	
Kamil & Rauscher (1990)					1a	A	
Kassem (1990)					1b	A, B	
Kenny (1975)					1b	A, B	
Knupfer(1993)					2	A	
Krieder (1992)					1b	A, B	
Lawrenz (1985)/Lawrenz & Munch (1984)					2	A, B	
Macdonald et al. (1966)					1a	A, B, C	
Marita (1965)					1a	A, B	
McHugh (1959)					1a	A	
Mehta (1993)					1b	A	
Merritt (1972)					1a	A, B, C	
Mevarech (1985)					1b	A	
Mevarech (1991)					1b	A	
Mevarech & Susak (1993)					1b	A	
Monroe (1922)			x				I
Moody & Gifford (1990)					2	A	
Mortlock (1969)			x		1a	A	
Park (1993)					1a	A, B	
Peterson & Janicki (1979)					1b	A, B	

Peterson et al. (1981)				1b	A, B	
Petty (1953a, 1953b)				1a	A	
Putbrese (1971/1972)		x	x	1a	A	
Sandby-Thomas (1983)				1a	A	
Shields (1927)		x	x			I
Slavin (1978)				1b	A, B, C	
Slavin & Karweit (1984a)				1b	A	
Slavin & Karweit (1984b, 1985)	x	x	x			I
Smith (1960/1961)	x	x	x	1a	A	
Spence (1958)	x	x	x			I
Stern (1971/1972)	x			1a	A, B	
Terwel et al. (1994)				1b	A	
Tingle & Good (1990)				1b	A	
Wallen & Vowels (1960)	x	x	x	1a	A	
Watson (1988)				1b	A	
Webb (1982a)				2	A	
Webb (1982b)				2	A	
Webb (1984)				2	A	
Webb et al. (1990)				1b	A	
Wright & Cowen (1985)				1b	A, B, C	
Yager et al. (1985)				1b	A	
Yueh & Alessi (1988)				2	A	
Ziegler (1981)				1b	A, B	
Zisk (1993)				1b	C	

Note. For analysis: 1a = homogeneous ability grouping versus no grouping, 1b = heterogeneous ability grouping versus no grouping, 2 = homogeneous ability grouping versus heterogeneous ability grouping. For outcome: A = achievement, B = attitudes, C = self-concept. For reason for excluding: I = group size > 10, II = within-group design, III = study confounds group composition with other factors.

*Only one type of analysis (namely, homogeneous ability grouping versus no grouping) and only one type of outcome (namely, student achievement) were included in Slavin's (1987, 1990) and Kulik and Kulik's (1987, 1991) meta-analyses on within-class grouping.

TABLE 2
Study features coded

Study feature	Description
	Outcome features
Measure source	Was the outcome measure standardized, researcher made, or teacher made?
Measure type	What type of outcome was measured?
Contrast type	Which of three questions about grouping was explored: homogeneous grouping versus no grouping, heterogeneous grouping versus no grouping, or homogeneous grouping versus heterogeneous grouping?
	Methodological features
Student equivalence	What attempts were made to achieve the equivalence of students in experimental and control conditions?
Teacher equivalence	What attempts were made to achieve the equivalence of teachers in experimental and control conditions?
School equivalence	What attempts were made to achieve the equivalence of schools in experimental and control conditions?
Overall design quality	Using a composite of student equivalence, teacher equivalence, and school equivalence, what is the overall design quality?
Teacher training equivalence	Did experimental teachers receive more or different training than control teachers?
Materials equivalence	Did experimental groups receive more or different material than control groups?
Rewards/grades equivalence	Did the experimental groups experience different rewards or grade structures than the control groups?
Overall instructional equivalence	Using a composite of teacher training equivalence, materials equivalence, and rewards/grades equivalence, what is the overall instructional equivalence?
Publication bias	Was the study published or unpublished?
	Substantive features
	Grouping characteristics
Grouping basis	What method(s) of assessment was used to group students?
Grouping specificity	Was grouping based on specific or general measures?
Group size	What was the average number of students in groups?
Group stability	Did group members stay together throughout the implementation?
	Instructor characteristics
Experimental teacher training	What was the amount of training (or experience in the strategy) given to the experimental teachers?
	Instructional characteristics
Type of small-group instruction	Was a cooperative learning strategy used in the experimental condition?
Control method of instruction	What instructional method was used in the control condition?

TABLE 2 (*continued*)

Study feature	Description
Goal structure of control condition	What was the goal structure (competitive, individualistic) used in the control condition?
	Student characteristics
Relative ability of students	What was the relative ability level of students in the class?
	Setting
Subject area	What was the subject area studied by the students?
Grade level	What was the students' grade level?
Class size	What was the average class size?
	Scope
Duration of treatment	What was the length of the experimental treatment?
Intensity of treatment	What was the intensity of the experimental treatment (e.g., hours per week)?

Note. The following study features were also coded but dropped from analyses due to 90% missing data or almost no variability: whose outcome was measured, outcome measure timing, type of experimental design, student expectations about grouping, teacher expectations about grouping, teaching experience, teacher gender, experimental method of instruction, medium of instruction, absolute ability level of the class, homogeneity of class ability, student race, student socioeconomic status, student gender, student attitudes toward subject, student attitudes toward experimental method, student self-concept, student locus of control, and country of implementation.

In two studies that used the Metropolitan Achievement Test and reported only grade equivalence data, we transformed the grade equivalence using a regression equation computed from technical information about the test (Hildreth, 1948). We also used means and SDs of the test norms to estimate the control group data for one study and the PSD for another study.

In studies that compared homogeneous grouping and heterogeneous grouping, the control group was the heterogeneous group. There were two reasons for this choice. First, the whole-class situation, which is a typical control condition, is usually heterogeneous. Second, meta-analyses on between-class ability grouping (e.g., Kulik & Kulik, 1987; Slavin, 1987) used the heterogeneous class as the control group.

Effect sizes from data in such forms as *t* tests, *F* tests, *p* levels, frequencies, and *r* values were computed via conversion formulas provided by Glass et al. (1981) and Hedges, Shymansky, and Woodworth (1989). When results were not significant, studies occasionally reported only a significance level. When the direction of the effect was not available, we estimated the effect size to be zero. For example, when the direction was reported, we used a "midpoint" approach (Sedlmeier & Gigerenzer, 1989) to estimate a representative *t* value (i.e., midpoint between 0 and the critical *t* value for the sample size to be significant). Statistical tests were performed on preliminary data sets to check whether estimated and unestimated effect sizes might be different. The nonsignificance of the test results confirmed that this effect size estimating procedure was reasonable.

Formulas for calculating effect sizes were entered into the Excel (Microsoft Corporation, 1993) computer program. Raw data for each finding were extracted by two coders separately and then checked for reliability. The initial coding agreement between the two coders was 65.51%. Disagreements between the two coders were subsequently resolved through discussion and further review of the disputed study finding.

Data Analyses

Data screening was performed using the SPSS (1994) frequency and descriptive procedures. Several study features with over 90% missing data or almost no variability (e.g., almost all teachers had over 2 years of experience) were dropped from further analysis. Categories within some variables (e.g., group size, subject area, grade) were combined based on frequency distributions and the preliminary results from the homogeneity analyses.

The homogeneity tests (Hedges & Olkin, 1985) were performed using DSTAT (B. T. Johnson, 1989), a meta-analysis computer program designed to integrate findings and analyze their variability. Effect sizes extracted from studies were corrected for bias and weighted by their sample sizes via formulas provided by Hedges and Olkin (1985). This approach not only provided a sample size weighted estimate of the overall effect of within-class grouping, but also allowed testing for heterogeneity in the aggregated effect sizes. When the homogeneity of the effect sizes was rejected (i.e., a significant chi-square value was obtained for Q_T), further exploration of the findings was done through the analysis of study features. A significant Q_B indicated that the study feature significantly moderated the magnitude of the effect sizes. When there were more than two levels of a study feature, Scheffé post hoc comparisons were performed to test for significant differences between levels. A significant Q_W for a level of the study feature indicated that the subset of effects sizes was heterogeneous.

Two major sets of analyses were performed. Analysis 1 involved effect sizes from studies that compared the effects of grouping with those of no grouping. Analysis 2 involved studies that directly investigated the effect of group ability composition on student achievement; no other outcomes could be analyzed. Effect sizes from these studies were based on direct comparisons of homogeneous ability grouping (i.e., students with a narrow range of relative abilities grouped together) with heterogeneous ability grouping (i.e., students with a wide range of relative abilities grouped together). Only three findings were extracted from studies in which method of instruction was confounded with group composition; these confounded effects were dropped from Analysis 2. Study features analyses were conducted for factors where sufficient variability existed.

Results

The outcome categories and analysis types extracted from each study are listed in Table 1. The number of independent findings and number of studies analyzed for each analysis type and outcome category are presented in Table 3.

TABLE 3
Maximum number of findings and studies analyzed per analysis for each category of outcomes

	Type of analysis	
Outcome	Grouping vs. no grouping	Homogeneous grouping vs. heterogeneous grouping
Achievement	103 (51)	20 (12)
Attitudes	30 (21)	
Self-concept	12 (10)	

Note. Values in parentheses are the numbers of studies from which the findings were extracted. Some studies provided outcomes in more than one category.

Analysis 1: Grouping Versus No Grouping

For Analysis 1, three major categories of outcomes (achievement, attitudes, and self-concept) were examined separately for the effect of within-class grouping. Full study feature analyses were performed on the achievement data only. Insufficient data were available for such analyses on the other outcomes.

Overall Effects of Within-Class Grouping on Student Achievement

The overall effect of within-class grouping on student achievement was based on 103 independent effect sizes extracted from 51 studies involving a total of 16,073 students. Student achievement measures included general ability tests, standardized subject matter achievement tests, and locally developed or teacher-made achievement tests. The mean sample size weighted effect size ($d+$) for within-class grouping was +0.17, which was significantly different from zero (95% confidence interval is +0.16 to +0.23). The mean effect size for the study findings randomly excluded from the analysis to avoid dependence problems was computed ($d+ = +0.20$, $n = 55$) and was not significantly different from the effects included.

On average, students learning in small groups within classrooms achieved significantly more than students not learning in small groups. In general, average students (i.e., those at the 50th percentile) in small-group classrooms performed at slightly above average (i.e., at about the 57th percentile) compared to students learning in classrooms without grouping. However, the homogeneity test showed that the effect sizes were heterogeneous, $Q_T(103) = 431.62$, $p < .05$. Examination of the individual findings revealed 74 effect sizes above zero, 5 equal to zero, and 24 effect sizes below zero. The range of effect sizes was from a low of −1.96 to a high of +1.52. Therefore, further exploration of the findings through the analysis of study features was warranted and is presented below.

What Factors Moderate the Effect of Within-Class Grouping on Student Achievement?

Study features analyses were performed on the 103 findings comparing the effects of homogeneous or heterogeneous grouping versus no grouping on achieve-

TABLE 4
Outcome features analyses: Grouping versus no grouping findings

Variables	Q_B	*n*	*d*+	95% CI	Q_W
Measure source	70.02*	98			
Standardized tests		61	+0.07	+0.03 to +0.11	134.92*
Teacher-made tests		16	+0.42	+0.26 to +0.59	60.56*
Researcher-made tests		21	+0.34	+0.28 to +0.39	145.78*
Measure type	59.24*	103			
Not geared to instruction		56	+0.08	+0.04 to +0.12	137.83*
Geared to instruction		47	+0.34	+0.29 to +0.39	234.55*
Contrast type	0.94	103			

* $p < .05$

ment. Results of the analyses are described below in three major categories: outcome features, methodological features, and substantive features.

Outcome features. Table 4 presents the results of the homogeneity tests on several outcome features: measure source, measure type, and contrast type. Measure source was a significant predictor, $Q_B(2) = 70.02$, $p < .05$. Post hoc analyses indicated that the mean weighted effect sizes for researcher-made tests (*d*+ = +0.34) and teacher-made tests (*d*+ = +0.42) were both significantly higher than for standardized tests (*d*+ = +0.07). Measure type was also significant. The effect sizes were larger when outcome measures were specifically geared to instructional content (*d*+ = +0.34) than when the outcome measures were not geared to what was taught (*d*+ = +0.08).

No significant effect was found for contrast type. The mean effect of homogeneous grouping compared to no grouping (*d*+ = +0.16) was similar to the mean effect of heterogeneous grouping versus no grouping (*d*+ = +0.19). Both mean weighted effect sizes were significantly positive, yet each set of findings was not uniform. A significant amount of within-group variability remained within each type of grouping; for homogeneous grouping, $Q_w = 246.51$, $p < .05$, and for heterogeneous grouping, $Q_w = 184.17$, $p < .05$.

Methodological features. Analyses of nine methodological features are presented in Table 5. Overall design quality (a composite of student equivalence, teacher equivalence, and school equivalence between experimental and control conditions) was not significantly related to the variability in effect sizes, $Q_B(1) = 3.38$, $p > .05$. Studies using either experimental control or statistical control showed significant positive effects for within-class grouping. However, overall instructional equivalence (a composite of teacher training equivalence, material equivalence, and reward equivalence between experimental and control conditions) was significantly related to the magnitude of the effect sizes. Greater achievement gains for within-class grouping occurred in studies that provided different instruction to the experimental condition (*d*+ = +0.25) than in studies that provided the same treatment (*d*+ = +0.02). Each of the three composite factors was significantly related to effect size variability. The effect of within-class grouping was higher when teachers in the grouped condition received more or different training as compared to those in the ungrouped condition (*d*+ = +0.42) than when teacher training was the same across conditions (*d*+ = +0.08). The within-class

TABLE 5
Methodological features analyses: Grouping versus no grouping findings

Variables	Q_B	*n*	*d*+	95% CI	Q_W
Student equivalence	5.83	103			
Teacher equivalence	0.13	99			
School equivalence	18.04*	86			
Experimental control		73	+0.12	+0.08 to +0.17	168.75*
Statistical control		13	+0.27	+0.22 to +0.33	196.96*
Overall design quality	3.38	103			
Teacher training equivalence	74.74*	56			
More or different		11	+0.42	+0.36 to +0.49	92.04*
Not different		45	+0.08	+0.04 to +0.13	138.73*
Materials equivalence	19.11*	69			
More or different		23	+0.26	+0.21 to +0.31	224.30*
Not different		76	+0.14	+0.10 to +0.18	188.70*
Rewards/grades equivalence	11.59*	99			
More or different		27	+0.29	+0.22 to +0.37	68.83*
Not different		51	+0.11	+0.06 to +0.15	154.39*
Overall instructional equivalence	18.78*	39			
Different treatment		23	+0.25	+0.19 to +0.31	83.89*
Not different treatment		16	+0.02	-0.07 to +0.11	27.68*
Publication bias	0.78	103			

* $p < .05$

grouping effect was higher when grouped classes employed more or different materials as compared to ungrouped classes (*d*+ = +0.26) than when the same materials were used across the two conditions (*d*+ = +0.14). Similarly, the within-class grouping effect was higher when grouped classes implemented more or different reward strategies as compared to ungrouped classes (*d*+ = +0.29) than when the same reward strategies were utilized across the two conditions (*d*+ = +0.11). Finally, whether findings were from published journals or from unpublished dissertations or reports was not related to the overall effect of within-class grouping.

Substantive features. Several groups of substantive features were analyzed. These were grouping characteristics, instructor characteristics, instructional methods, student characteristics, setting factors, and scope of the treatment.

Table 6 presents the results of the study features analyses on *grouping characteristics.* First, grouping basis was significantly related to the variability in effect sizes. Post hoc comparisons indicated that the mean weighted effect size was significantly higher when grouping was based on mixed sources (*d*+ = +0.33) than on either standardized tests alone (*d*+ = +0.10) or teacher-made tests or teacher judgement alone (*d*+ = –0.02). Second, grouping specificity significantly explained variability in the effect sizes. The mean effect size was highest when grouping was based on assessment of specific or general ability plus other factors (*d*+ = +0.39). Third, group size was significantly related to the magnitude of the effect sizes; the average effect size for 3–4 member groups (*d*+ = +0.22) was significantly higher than that for 5–7 member groups (*d*+ = –0.02). Finally, group stability did not significantly explain the variability in the effect sizes.

Study features analyses on *instructor and instructional characteristics* are

TABLE 6
Substantive features analyses (grouping characteristics): Grouping versus no grouping

Variables	Q_B	*n*	*d*+	95% CI	Q_W
Grouping basis	68.84*	81			
Standardized tests		48	+0.10	+0.03 to +0.17	97.04*
Teacher tests or judgment		12	–0.02	–0.09 to +0.05	65.11*
Mixed sources		21	+0.33	+0.28 to +0.38	127.63*
Grouping specificity	18.12*	84			
Ability specific to subject		58	+0.19	+0.15 to +0.24	297.79*
General ability		9	+0.19	–0.02 to +0.40	13.79
Both specific & general ability		8	+0.10	+0.02 to +0.18	15.57*
Specific/general ability + other		9	+0.39	+0.28 to +0.49	18.62
Group size	31.99*	92			
Pairs		13	+0.15	+0.06 to +0.24	34.94*
Small (3–4)		38	+0.22	+0.22 to +0.16	132.30*
Medium (5–7)		17	–0.02	–0.02 to –0.09	75.28*
Large (8–10)		24	+0.11	+0.01 to +0.21	32.78*
Group stability	0.14	40			

* $p < .05$

presented in Table 7. Teacher experience with, or amount of training received for, the experimental treatment was related to the magnitude of the effect sizes. The average effect size was +0.17 for no experience or training, +0.24 for information only, +0.31 for minimal training or experience, and +0.57 for extensive training or experience. All four effect sizes were significantly different from zero. The average effect size for moderate experience or training was +0.02, which was not significantly different from zero.

TABLE 7
Substantive features analyses (instructor and instructional characteristics): Grouping versus no grouping findings

Variables	Q_B	*n*	*d*+	95% CI	Q_W
Experimental teacher training	55.72*	61			
None		2	+0.17	+0.06 to +0.28	0.25
Info. only		9	+0.24	+0.15 to +0.34	22.13*
Minimal		36	+0.31	+0.25 to +0.36	160.74*
Moderate		11	+0.02	–0.05 to +0.08	74.42*
Extensive		3	+0.57	+0.32 to +0.82	0.23
Type of small-group instruction	10.27*	103			
Cooperative learning		25	+0.28	+0.21 to +0.34	57.63*
Other		77	+0.15	+0.11 to +0.19	360.70*
Control method of instruction	56.46*	102			
Traditional		77	+0.24	+0.20 to +0.28	314.36*
Individualized mastery learning		15	+0.15	+0.07 to +0.24	20.88
Other		10	–0.12	–0.20 to –0.03	36.90*
Goal structure of control condition	0.51	67			

* $p < .05$

The experimental method of instruction explained a significant amount of variability in effect sizes, $Q_B(9) = 42.96$, $p < .05$, but there were too few cases to reliably examine most methods. Consequently, we divided the studies into two types of small-group instruction: cooperative learning and other. Type of instruction was significantly related to the size of the treatment effect. Small groups using cooperative learning with outcome interdependence ($d+ = +0.28$) achieved significantly higher than other small groups without outcome interdependence ($d+ = +0.15$).

The method of instruction used in the control condition was also significantly related to the size of the grouping effect. The effect of placing students in small groups was superior when the control method of instruction was either traditional, frontal teaching ($d+ = +0.24$) or individualized mastery learning ($d+ = +0.15$). However, when within-class grouping was compared to other methods of instruction such as experiential learning, within-class grouping was not superior ($d+ = -0.12$). Finally, the goal structure (competitive, individualistic) used in the control condition was not related to effect size variability.

Table 8 presents the results of the study features analyses of *student characteristics, setting, and scope of the treatment.* The relative ability of students was significantly related to the within-class grouping effect sizes, $Q_B(2) = 6.20$, $p < .05$. While low-ability students, medium-ability students, and high-ability students all benefitted from being placed in small groups ($d+ = +0.37$, $+0.19$, and

TABLE 8
Substantive features analyses (student characteristics, setting, and scope): Grouping versus no grouping findings

Variables	Q_B	n	$d+$	95% CI	Q_W
Relative ability of students	6.20*	53			
Low		24	+0.37	+0.30 to +0.44	104.44*
Medium		11	+0.19	+0.06 to +0.32	26.35*
High		18	+0.28	+0.15 to +0.41	47.07*
Subject area	4.52*	103			
Math/science		65	+0.20	+0.16 to +0.24	316.02*
Reading/language arts/other		38	+0.13	+0.09 to +0.18	111.08*
Grade level	27.55*	98			
Early elementary (1–3)		30	+0.08	+0.02 to +0.14	157.81*
Late elementary (4–6)		36	+0.29	+0.24 to +0.35	132.05*
Secondary (7–12)		25	+0.17	+0.12 to +0.23	85.98*
Postsecondary		7	+0.19	–0.03 to +0.42	8.55
Class size	46.70*	103			
Small (less than 25)		33	+0.22	+0.17 to +0.27	114.80*
Medium (26–35)		50	+0.06	+0.01 to +0.11	178.46*
Large (more than 35)		20	+0.35	+0.28 to +0.42	91.66*
Duration of treatment	0.34	103			
Intensity of treatment	14.49*	103			
Low (≤ 1 period/week)		40	+0.08	+0.03 to +0.14	141.83*
High (> 1 period/week)		63	+0.22	+0.18 to +0.25	275.30*

* $p < .05$

+0.28, respectively), the effects were not uniform. In particular, low-ability students achieved significantly more than medium-ability students.

Within-class grouping positively affected student learning in all subject areas. However, the effect of within-class grouping was significantly larger in math and science (*d*+ = +0.20) than in reading, language arts, and other courses (*d*+ = +0.13). We were able to identify studies on within-class grouping at grade levels ranging from first grade to college, but the number of studies at any one level was small, and the findings appeared variable. The largest effect sizes were obtained in the late elementary grades (Grades 4–6; *d*+ = +0.29). Within-class grouping had a significantly positive effect for small (less than 25), medium (26–35), and large (more than 35) classes. The grouping effect was stronger in larger classes (*d*+ = +0.35) than in small (*d*+ = +0.22) or medium (*d*+ = +0.06) classes.

Whether the experimental treatment was brief (less than 4 weeks), medium (4 to 16 weeks), or long (more than 17 weeks) was not significantly related to the size of the effect. However, treatment intensity seemed to be a significant moderator of the grouping effect. The effect was larger in studies of high treatment intensity (more than 1 period per week, *d*+ = +0.22) than in studies of low treatment intensity (less than or up to 1 period per week, *d*+ = +0.08).

Overall Effects of Within-Class Grouping on Other Student Outcomes

The literature included a sufficient number of studies to allow a rudimentary exploration of the relationship between the use of within-class grouping and outcomes other than achievement. The mean effect sizes for student attitudes and self-concept are reported below.

Student attitudes. Thirty independent findings extracted from 21 studies compared student attitudes from within-class groups with student attitudes in no-grouping conditions. Measures of student attitudes included attitudes toward the subject matter, the instructional approach, and others (peers, school, etc.). Results of the homogeneity analyses on student attitudes are presented in Table 9. Overall, within-class grouping was positively related to student attitudes (*d*+ = +0.18, 95% confidence interval is +0.13 to + 0.23). In particular, students in the grouped classes had significantly more positive attitudes toward the subject matter concerned (*d*+ = +0.18). However, attitudes toward the instructional approach were

TABLE 9
Overall effects of within-class grouping on other student outcomes: Grouping versus no grouping findings

Outcomes	*n*	*d*+	95% CI	Q_T
Overall student attitudes	30	+0.18	+0.13 to +0.23	327.04*
Attitudes toward subject	16	+0.18	+0.13 to +0.24	280.76*
Attitudes toward instructional approach	5	-0.13	-0.33 to +0.06	18.39*
Other attitudes	9	+0.26	+0.14 to +0.37	16.15
Overall student self-concept	12	+0.09	-0.00 to +0.19	11.11
General self-concept	6	+0.16	+0.02 to +0.31	3.33
Academic self-concept	6	+0.04	-0.09 to +0.16	6.15

* $p < .05$

not significantly different between students in the grouped and ungrouped classes.

Student self-concept. Ten studies investigated the effects of within-class grouping on student self-concept. From those studies, 12 independent findings were extracted and analyzed (see Table 9). Measures of self-concept included general self-concept and domain-specific academic self-concept. Overall, the mean weighted effect size for student self-concept was +0.09, which is not significantly different from zero (95% confidence interval is –0.00 to + 0.19). However, students in grouped classes had significantly higher general self-concept than students in the ungrouped classes ($d+ = +0.16$). Domain-specific academic self-concept was not significantly different between the grouped and ungrouped classes.

Analysis 2: Homogeneous Grouping Versus Heterogeneous Grouping

Twenty findings from 12 studies directly compared the effects of homogeneous ability grouping with heterogeneous ability grouping. All the studies employed interactive small-group learning. Methods of instruction were held constant across the two conditions. Most studies employed experimental designs and implemented equivalent instructional treatments in terms of teacher training, reward strategies, and curriculum materials across the two conditions.

Overall Effect of Group Ability Composition on Student Achievement

The weighted mean effect size for group ability composition was +0.12, which is significantly different from zero (95% confidence interval is +0.01 to +0.24). Thus, the result of aggregating the 20 independent effect sizes indicated a slight superiority of homogeneous ability groups over heterogeneous ability groups in promoting student achievement. However, the slight superiority of homogeneous ability grouping was not uniform across findings, $Q_T(19) = 43.90$, $p < .05$. The effect sizes ranged from –1.75 to +1.12, with 13 effect sizes above zero favoring homogeneous ability groups, 1 effect size equal to zero, and 6 effect sizes below zero favoring heterogeneous ability groups. The heterogeneity of the effect sizes, therefore, warranted further analyses, which are presented in Table 10.

Moderators of the effects of group ability composition. There were sufficient differences in the design of the group composition studies to explore three study

TABLE 10
Study features analyses: Homogeneous ability grouping versus heterogeneous ability grouping findings

Variables	Q_B	n	$d+$	95% CI	Q_W
Type of small-group instruction	1.48	20			
Relative ability of students	11.42*	13			
Low		4	-0.60	-1.11 to -0.09	7.27
Medium		4	+0.51	+0.11 to +0.90	1.72
High		5	+0.09	-0.25 to +0.42	0.56
Subject area	8.61*	20			
Math/science		16	-0.00	-0.15 to +0.14	25.74
Reading		4	+0.36	+0.16 to +0.55	9.55*

* $p < .05$

features. One of these, type of small-group instruction, did not explain variability in the effect sizes. The small advantage of homogeneous grouping compared to heterogeneous grouping was not modified by whether students learned cooperatively with outcome interdependence or students learned in small groups without outcome interdependence.

However, the effects of group ability composition were different for students of different relative ability. While low-ability students learned significantly more in heterogeneous ability groups than in homogeneous ability groups ($d+ = -0.60$), medium-ability students benefitted significantly more in homogeneous ability groups than in heterogeneous ability groups ($d+ = +0.51$). For high-ability students, group ability composition made no significant difference ($d+ = +0.09$). Furthermore, the Q_Ws for the three mean effect sizes reported were all homogeneous, which indicated consistent and meaningful findings within each subset of results.

Subject area of instruction was also a significant moderator of the effect of group ability composition. In math and science, the effect of group ability composition was not significantly different from zero; overall, homogeneous groups performed as well as heterogeneous groups. However, four effect sizes compared homogeneous ability groups with heterogeneous ability groups in reading, and the mean effect size ($d+ = +.36$) revealed that, on average, homogeneous ability groups learned significantly more than heterogeneous ability groups in reading.

Discussion

This study quantitatively synthesized the literature on the effects of within-class grouping on student achievement, attitudes, and self-concept. The results complement and greatly extend the findings reported in other research integrations (Kulik & Kulik 1987, 1991; Slavin, 1987). Each of the reviews, including this one, reported positive effects of within-class grouping. Our review included far more studies, evaluated student learning and other outcomes, explored variability in study findings, and examined studies that directly compared types of small-group composition.

The results of this meta-analysis suggest that there are small but positive effects of placing students in groups within the classroom for learning. On average, students placed in small groups achieved more, held more positive attitudes, and reported higher general self-concept than students in nongrouped classes. However, the magnitude of the effect sizes varied across findings. Several study features significantly moderated the effect of within-class grouping on student achievement. These study features included outcome measure source and type, instructional treatment equivalence of the control and experimental groups, type of small-group instruction and amount of teacher training in the experimental condition, type of instructional method in the control condition, grouping basis and specificity, group size, group ability composition, relative ability of students, subject area, grade level, class size, and intensity of treatment.

Outcome Measure Source and Type

Measuring achievement by locally developed tests or by standardized tests can produce significantly different treatment effects. This phenomenon has been noticed by several meta-analysts. For example, Rosenshine and Meister (1994) in their review of reciprocal teaching found a significantly higher effect size for experimenter-made tests than standardized tests. One explanation Rosenshine and Meister offered was that experimenter-made tests were generally easier to answer, or more instructionally sensitive, than standardized tests. The topics in the experimenter-made tests were discussed during instruction, whereas those in the standardized tests were unfamiliar to the students.

In this meta-analysis, we found that the effect of small-group learning was much higher when achievement was measured with teacher-made tests than when researcher-made tests were used. Achievement measured with researcher-made tests was, in turn, higher than that measured with standardized tests. Similarly, we also showed that the effect sizes were higher when the outcome measures were geared to instruction than when they were not geared to instruction. Therefore, one explanation for the difference between locally developed tests and standardized tests is that teacher-made tests may have a closer match with local instructional objectives than researcher-made tests or standardized tests. Similarly, researcher-made tests may have a closer match to the local instructional objectives than standardized tests. Thus, locally made tests may reflect the large influence of within-class grouping on proximal instructional objectives, while standardized tests may reflect the small influence of within-class grouping on distal instructional objectives.

Another possible explanation is that teacher-made and researcher-made tests may be biased in favor of the experimental group. If the goal of the research is to measure acquisition of local instructional objectives, care should be taken to ensure that the tests measure what has been taught in both experimental and control conditions.

Instructional Treatment and Teacher Training

Differential instructional treatments can significantly moderate the effect of within-class grouping. Ungrouped classes usually employed a single set of materials for all the students in the class, but that did not always occur in grouped classes. In some studies, teachers employed the same set of materials for all the groups as well as the control classes; in others, teachers employed different materials across groups. The effect sizes were higher when instructional materials were varied for different groups than when the same set of materials was used for all the students. One argument for using small groups is to provide instruction which better meets the learning needs of individual students. Varying the instructional material is one means of providing more adaptive instruction (Kulik & Kulik, 1991).

The type of small-group instruction used in the experimental classes can moderate the achievement effects of within-class grouping. Often in cooperative small-group instruction, group members are positively interdependent for their learning outcomes. Each member contributes to the overall group goal and receives some form of group reward. The results of this meta-analysis suggest that small groups appeared to learn more when there was outcome interdependence among the group members.

Amount of training given to experimental and control teachers can significantly moderate the effect of within-class grouping. The largest effect sizes occurred in studies where the experimental teachers received much more or very different training than the control teachers. Even so, when teacher training was held constant across the control and experimental teachers, there was still a small positive effect for within-class grouping.

One anomaly to the linear effect of training occurred for teachers with moderate amounts of training. One tentative explanation is that teachers with little training adapt the new method to their existing practices and teaching philosophy, while teachers with extensive training adapt their practices and teaching philosophy to the new method. Teachers with moderate training may make neither adaptation successfully, which would minimize the benefits of small groups.

There were also variations in the instructional methods used in the ungrouped control classes. Within-class grouping had a significantly positive effect on student learning when compared with traditional whole-class instruction and individual seatwork or individualized mastery learning. However, when within-class grouping was compared with other methods of instruction (e.g., experiential whole-class learning), it was not always superior.

Overall, it appears that the positive effects of within-class grouping are maximized when the physical placement of students into groups for learning is accompanied by modifications to teaching methods and instructional materials. Merely placing students together is not sufficient for promoting substantive gains in achievement. Consequently, Slavin's (1987) practice of discarding studies where grouping method is confounded with other factors should be viewed with caution, since it may well remove not only instructionally relevant studies but those with the largest effects.

Grouping Strategies: Grouping Basis, Group Size, and Group Stability

Teachers employed a variety of strategies in grouping students for learning. Some groups were formed using standardized test results; others were formed using teacher-made test results or teachers' knowledge of students' abilities. Some group formations were based on general ability or specific ability in a subject matter, and others on a mixture of considerations. The basis and specificity on which small groups were formed significantly moderated the effect of within-class grouping on student achievement. Larger effects occurred when group formation was based on mixed sources and involved more considerations than ability alone. Just as the classroom or school is a social community, the small group where students learn together is also a small social community. Group interaction may be more positive when groups are cohesive, and teacher judgments of compatibility may help groups function well (Cohen, 1994).

Optimally sized groups for learning seem to be small, 3- to 4-member teams. While pairs achieved significantly more than students in ungrouped classes, the large, 6- to 10-member groups did not learn significantly more than students from ungrouped classes. Finally, whether groups stayed together or changed members over time was not significantly related to the effect of within-class grouping on student achievement.

Subject Area and Class Size

The magnitude of the within-class grouping effect varied according to subject area and class size. There were larger effects of within-class grouping in math and science than in reading, language arts, and other courses. This may be due to the different nature of the learning tasks involved in these subject areas. Tasks in math are usually more hierarchical (i.e., where the level and complexity of the material to be learned must be adjusted to suit differences in the prior knowledge of students). Specific assistance from peers may help students progress through such hierarchical learning faster.

Large classes seemed to benefit more from within-class grouping. Grouping may provide a means for more adjusted instruction to the students in these classes and possibly greater opportunities for peer interaction and more active engagement of each pupil.

Group Ability Composition and the Relative Ability of Students

Group ability composition had a differential effect on student learning. In general, homogeneous ability groups achieved more than heterogeneous ability groups in studies that directly compared them. However, the superiority of homogeneous ability composition is not uniform for students of different relative ability. While low-ability students learned significantly more in heterogeneous ability groups than in homogeneous ability groups, medium-ability students benefited significantly more in homogeneous ability groups than in heterogeneous ability groups. For high-ability students, group ability composition made no significant difference.

Several mediational mechanisms are plausible explanations for these findings. First, according to Webb (1982a, 1982b, 1984), learning in small groups depends on giving and receiving explanations. Giving explanations helps tutors clarify and organize their own learning better. Receiving elaborated explanations helps tutees correct misconceptions and learn appropriate learning strategies. Not receiving an explanation—that is, receiving no response at all or receiving simply the answer—reduces achievement. Low-ability students may gain most in heterogeneous groups from having other students provide them with timely and elaborated assistance and guidance. In contrast, when low-ability students are placed in homogeneous groups there may be no student capable of providing those explanations. High-ability students may benefit from being placed in heterogeneous groups to the extent that they are often called upon to provide elaborated explanations by their less able peers. Medium-ability students, however, may act neither as tutor nor tutee and, therefore, neither give nor receive explanations. Consequently, heterogeneous grouping is not as beneficial for these students. Homogeneous grouping may be better for medium-ability students because they may share in giving and receiving explanations among themselves.

Second, group cohesiveness can lead to increased performance by enhancing members' commitment to the group task (Mullen & Copper, 1994). In contrast to heterogeneous grouping, homogeneous grouping may be particularly conducive to group cohesiveness since students may share similar expectations about group goals. Medium-ability and high-ability students may especially benefit without compromising their aspirations or pace of learning to accommodate the lower-ability students.

Third, adaptation of instruction may be important in realizing the benefits of homogeneous grouping. Appropriate tailoring of instruction for low-ability students may place extraordinary demands on the teacher. Furthermore, it may well be the case that low-ability students placed in homogeneous groups may suffer if the demands for learning are set too low; if these students feel isolated, inadequate, or incompetent; or if the teacher has negative performance expectations. Medium-ability groups may require the least adaptation of regular teaching materials and therefore exhibit the previously mentioned benefits of homogeneous grouping. Finally, high-ability students in homogeneous groups may work through the regular material at a faster pace and may challenge one another to elaborate their learning further.

Overall, then, we found no evidence that one form of grouping was *uniformly* superior for promoting the achievement of all students. Low-ability students gained most from being placed in heterogeneous groups with students who might provide them with individual guidance and assistance. The assistance provided may, in turn, be beneficial for the tutors in helping them develop a deeper, more structured understanding of the material. In contrast, medium-ability students gained most from being placed in homogeneous groups. For these students, sharing in giving and receiving explanations, high group cohesiveness, and appropriate instructional materials may be important factors and should be explored further.

Speculations, Limitations, and Future Directions

The myriad of factors which may distinguish among classroom grouping practices eventually come to influence one of two interrelated processes: students' motivation to learn and students' processing, acquisition, and retention of information. Exploring these factors helps to understand the superiority of an instructional practice and the conditions for optimal implementation. But grouping practices themselves do not directly affect motivation and learning; it depends on how they are used.

For example, it is generally accepted that teacher, peer, and personal expectations influence student goal setting and motivation to learn. Teachers using whole-class instruction may negatively influence student expectations if they selectively encourage students to excel, give students unequal opportunities for responding, and praise (or criticize) students differentially when they are correct (or incorrect) according to ability (Cooper & Good, 1983). But if teachers provide equitable opportunities and proportionally uniform praise or otherwise exhibit facilitative behaviors and attitudes, then the expectations of most students may be raised. Similarly, teachers using small-group instruction may positively influence student expectations if they assign group tasks that depend on the contributions of all students (Abrami et al., 1995). But if teachers assign group tasks that depend on the contribution of only the brightest students, the individual expectations for learning of most students may suffer.

Similarly, peer influences may either facilitate or discourage student performance. Students may establish a culture that promotes academic goals and achievement. In contrast, students may challenge teachers, obstruct academic activity, and misuse educational resources. These deleterious effects may be more likely in whole-class instruction when relatively weak students defend themselves against public academic humiliation and failure, but they are not unheard of in small groups (e.g., Salomon & Globerson, 1989). Abrami, Chambers, d'Apollonia, Farrell, and De Simone (1992) found negative effects of heterogeneous small-group learning for low-ability and learned helpless students whose groups had not succeeded at the learning task. They suggested that both self-deprecation and blame from teammates combine to provide a double disincentive for these students. In contrast, homogeneously grouped low achievers may suffer from a lack of appropriately behaving role models, which may increase the likelihood that they will mimic one another's off-task behaviors (Felmlee & Eder, 1983).

In undertaking this review, we attempted to explore the instructional and learning processes which distinguish whole-class instruction from small-group instruction and among heterogeneous and homogeneous small groups. However, the complexity of these processes and the paucity of evidence limited the extent to which we were successful. Now that we have moved closer to determining whether within-class grouping is effective, when it is effective, and with whom it is effective, it is time to devote greater energy to understanding why it is effective. We hope this review sets the stage for such inquiry.

We caution the reader that this meta-analysis, like others, does not allow one to make strong causal inferences, particularly with regard to explanatory features. Not only were we unable to extract information from every study about the existence of particular factors, which reduces the sensitivity of the analyses, but the study features were often intercorrelated while the heterogeneity within categories of study features were not resolved in many cases, which makes unambiguous interpretation impossible and untempered conclusions unwise.

It is possible that factors not identified by us or the primary researchers may explain some of the variability in study findings. It is also possible that some of the variability may be explained by substantive and methodological features which correlate or interact with those we identified. For example, in those studies comparing grouping with no grouping, the superiority of cooperative learning over other methods of instruction may be explained or exaggerated by the extra, recent teacher training often involved in undertaking the former method. Contrariwise, it may be that differences in classroom experience with cooperative versus other methods attenuate the size of the relationship with study findings. Such explanations could not be adequately tested by us; they await verification through additional primary research.

Conclusions and Recommendations

The practice of within-class grouping is supported by the results of this review. Within-class grouping appears to be a useful means to facilitate student learning, particularly in large classes and especially in math and science courses. Small teams of three to four members seem more effective than larger groups. Low-ability students benefit most when placed in mixed-ability groups, but medium-ability students benefit most in relatively homogenous ability groups. Cooperative learning with outcome interdependence helps facilitate small-group learning. Furthermore, teacher training in, and experience with, small-group instructional strategies helps maximize student learning. Finally, the best within-class grouping practices combine the physical placement of students into groups with the adaptation of instruction methods and materials for small-group learning.

Note

[1]Slavin (1990) also quantitatively reviewed the research on the effects of ability grouping on the achievement of secondary students. While the review focused on the effects of between-class grouping, Slavin also summarized the findings of several studies of within-class grouping, but without reporting effect sizes, and concluded that the effects of within-class grouping were largely nonsignificant.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Abrami, P. C., Chambers, B., d'Apollonia, S., Farrell, M., & De Simone, C. (1992). Group outcome: The relationship between group learning outcome, attributional style, academic achievement, and self-concept. *Contemporary Educational Psychology, 17,* 201–210.

Abrami, P. C., Chambers, B., Poulsen, C., De Simone, C., d'Apollonia, S., & Howden, J. (1995). *Classroom connections: Understanding and using cooperative learning.* Toronto, Ontario, Canada: Harcourt-Brace.

Abrami, P. C., Cohen, P. A., & d'Apollonia, S. (1988). Implementation problems in meta-analysis. *Review of Educational Research, 58,* 151–179.

Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology, 82,* 219–231.

*Abu, R. B. (1993). The effects of cooperative learning methods on achievement, retention, and attitudes of selected vocational home economics students (Doctoral dissertation, North Carolina State University, 1993). *Dissertation Abstracts International, 54,* 3350A.

Allan, S. D. (1991). Ability-grouping research reviews: What do they say about grouping and the gifted? *Educational Leadership, 48,* 60–74.

*Allen, W. H., & VanSickle, R. L. (1984). Learning teams and low achievers. *Social Education, 48,* 60–64.

*Amaria, R. P., Biran, L. A., & Leith, G. O. M. (1968). Individual versus cooperative learning: Part 1—Influence of intelligence and sex. *Educational Research, 2,* 95–103.

*Armstrong, N. (1993). The effects of cooperative learning on gifted students in heterogeneous and homogeneous groups (Doctoral dissertation, Ball State University, 1993). *Dissertation Abstracts International, 54,* 2457A.

*Ballman, T. L. (1988). Is group work better than individual work for learning Spanish?: The findings of one study. *Hispania, 71,* 180–185.

*Bejarano, Y. (1987). A cooperative small group methodology in the language classroom. *TESOL Quarterly, 21,* 483–504.

*Berge, Z. L. (1990). Effects of group size, gender, and ability grouping on learning science process skills using microcomputers. *Journal of Research in Science Teaching, 27,* 747–759.

Bierden, J. E. (1970). Behavioral objectives and flexible grouping in seventh grade mathematics. *Journal for Research in Mathematics Education, 1,* 207–217.

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

*Bright, G. W., Harvey, J. G., & Wheeler, M. M. (1980). Achievement grouping with mathematics concept and skill games. *Journal of Educational Research, 73,* 265–269.

*Campbell, A. L. (1965). A comparison of the effectiveness of two methods of class organization for the teaching of arithmetic in junior high school (Doctoral dissertation, Pennsylvania State University, 1964). *Dissertation Abstracts International, 26,* 813.

*Carter, G., & Jones, M. G. (1993, April). *The relationship between ability paired interactions and the development of fifth graders' concepts of balance.* Paper presented at the annual conference of the National Association for Research in Science Teaching, Atlanta, GA. (ERIC Document Reproduction Service No. ED 361 175)

*Chang, S. (1993). The effect of group reward on student motivation, interaction, emotion, and achievement in cooperative learning small groups (Doctoral dissertation, University of California, Los Angeles, 1993). *Dissertation Abstracts International, 55,* 906A.

*Cignetti, M. J. (1974). A study of intraclass grouping and traditional groupings on students' terminal achievements during the last nine weeks in first semester typewriting (Doctoral dissertation, University of Pittsburgh, 1974). *Dissertation Abstracts International, 35,* 5765A.

Cohen, E. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64,* 1–35.

Coldiron, J. R., Braddock, J. H., & McPartland, J. M. (1987, April). *A description of school structures and practices in elementary, middle, and secondary schools.* Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC.

Cooper, H. M., & Good, T. L. (1983). *Pygmalion grows up: Studies in the expectation communication process.* New York: Longman.

*DeVries, D. L., & Edwards, K. J. (1973). Learning games and student teams: Their effects on classroom process. *American Educational Research Journal, 10,* 307–318.

*Dewar, J. A. (1963). Grouping for arithmetic instruction. *Elementary School Journal, 63,* 266–269.

Eddleman, V. K. (1971). A comparison of the effectiveness of two methods of class organization for arithmetic instruction in grade five (Doctoral dissertation, Northeast Louisiana University, 1971). *Dissertation Abstracts International, 32,* 1744A.

*Evans, M. M. (1942). *The effect of variance grouping on reading achievement.* Unpublished doctoral dissertation, University of Pittsburgh.

*Fantuzzo, J. W., Polite, K., & Grayson, N. (1990). An evaluation of reciprocal peer tutoring across elementary school settings. *Journal of School Psychology, 28,* 309–323.

Felmlee, D., & Eder, D. (1983). Contextual effects in the classroom: The impact of ability groups on student attention. *Sociology of Education, 56,* 77–87.

Gage, N. L. (1959). Stanford achievement test. In O. K. Buros (Ed.), *The fifth mental measurements yearbook* (pp. 76–80). Highland Park: The Gryphan Press.

Gamoran, A. (1987). Organization, instruction, and the effects of ability grouping: Comment on Slavin's "best-evidence synthesis." *Review of Educational Research, 57,* 341–345.

Gleser, L. J., & Olkin, I. (1994). Stochastically dependent effect sizes. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 339–355). New York: Russell Sage Foundation.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Good, T. L., & Marshall, S. (1984). Do students learn more in heterogeneous or homogeneous groups? In P. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The Social context of instruction: Group organization and group processes* (pp. 15–38). Orlando, FL: Academic Press.

Hallinan, M. (1984). Summary and conclusions. In P. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction: Group organization and group processes* (pp. 229–240). Orlando, Fl.: Academic Press.

*Hallinan, M. T., & Sorensen, A. B. (1985). Ability grouping and student friendships. *American Educational Research Journal, 22,* 485–499.

*Harrah, D. D. (1956). A study of the effectiveness of five kinds of grouping in the classroom (Doctoral dissertation, University of Virginia, 1955). *Dissertation Abstracts International, 16,* 715.

*Hay, J. A. (1981). Effects of cooperative goal structuring on sixth grade science students' abilities to initiate task and maintenance group behaviours (Doctoral dissertation, Michigan State University, 1980). *Dissertation Abstracts International, 41,* 3857A.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1989). *A practical guide to modern methods of meta-analysis* (Report No. ISBN-0-87355-081-1). (ERIC Document Reproduction Service No. ED 309 952)

*Heller, L. R., & Fantuzzo, J. W. (1993). Reciprocal peer tutoring and parent partnership: Does parent involvement make a difference? *School Psychology Review, 22,* 517–534.

Hildreth, G. H. (1948). *Manual for interpreting Metropolitan Achievement Tests.* Yonkers, NY: World Book Company.

*Hudgins, B. B. (1960). Effects of group experience on individual problem solving. *Journal of Educational Psychology, 51,* 37–42.

*Hulten, B. H., & DeVries, D. L. (1976). *Team competition and group practice: Effects on student achievement and attitudes* (Rep. No. 212). (ERIC Document Reproduction Service No. ED 154 021)

Hunter, E. J., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

*Janicki, T. C., & Peterson, P. L. (1981). Aptitude treatment interaction effects of variations in direct instruction. *American Educational Research Journal, 18,* 63–82.

Johnson, B. T. (1989). DSTAT, software for the meta-analytic review of research literatures [Computer software]. Hillsdale, NJ: Erlbaum.

Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research.* Edina, MN: Interaction Book.

*Johnson, D. W., Johnson, R. T., & Skon, L. (1979). Student achievement on different types of tasks under cooperative, competitive, and individualistic conditions. *Contemporary Educational Psychology, 4,* 99–106.

*Jones, D. M. (1948). An experiment in adaptation to individual differences. *Journal of Educational Psychology, 39,* 257–273.

*Kamil, M. L., & Rauscher, W. C. (1990). Effects of grouping and difficulty of materials on reading achievement. *National Reading Conference Yearbook, 39,* 121–127.

*Kassem, C. L. (1990). *Cooperative versus individualistic learning strategies for undergraduate teaching: A comparative study.* Unpublished doctoral dissertation, Georgia State University, College of Education.

*Kenny, P. F. (1975). *Effects of group interaction stimulated by competition between groups as a motivating technique in a ninth grade mathematics classroom: Final report.* Washington, DC: National Institute of Education, Department of Health, Education, and Welfare, Office of Research, Basic Studies Division. (ERIC Document Reproduction Service No. ED 103 266)

*Knupfer, N. N. (1993). Logo and transfer of geometry knowledge: Evaluating the effects of student ability grouping. *School Science and Mathematics, 93,* 360–368.

*Kreider, P. S. (1992). Achievement in physical science using cooperative mastery learning (Doctoral dissertation, Florida Institute of Technology, 1992). *Dissertation Abstracts International, 53,* 3861A.

Kulik, J. A., & Kulik, C.-L. C. (1987). Effects of ability grouping on student achievement. *Equity and Excellence, 23,* 22–30.

Kulik, J. A., & Kulik, C.-L. C. (1991). *Research on ability grouping: Historical and contemporary perspectives.* Storrs: University of Connecticut, National Research Center on the Gifted and Talented. (ERIC Document Reproduction Service No. ED 350 777)

*Lawrenz, F. (1985). Aptitude treatment effects of laboratory grouping method for students of differing reasoning ability. *Journal of Research in Science Teaching, 22,* 279–287.

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

*Macdonald, J. B., Harris, T. L., & Mann, J. S. (1966). Individual versus group instruction in first grade reading. *The Reading Teacher, 19,* 646–652.

*Marita, S. M. (1965). *A comparative study of beginning reading achievement under three classroom organizational patterns: Modified individualized, three to five groups, and whole class language experience.* Milwaukee, WI: Marquette University. (ERIC Document Reproduction Service No. ED 003 477)

*McHugh, W. J. (1959). Team learning in skills subjects in intermediate grades. *Journal of Education, 142,* 22–51.

*Mehta, J. I. (1993). Cooperative learning in computer programming at the college level (Doctoral dissertation, University of Illinois at Chicago, 1993). *Dissertation Abstracts International, 54,* 1309A.

*Merritt, P. W. (1972). *The effects of variations in instruction and final unit evaluation procedures on community college beginning algebra classes.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

*Mevarech, Z. R. (1985). The effects of cooperative mastery learning strategies on mathematical achievement. *Journal of Educational Research, 78,* 372–377.

*Mevarech, Z. R. (1991). Learning mathematics in different mastery environments. *Journal of Educational Research, 84,* 225–231.

*Mevarech, Z. R., & Susak, Z. (1993). Effects of learning with cooperative mastery method on elementary students. *Journal of Educational Research, 86,* 197–205.

Microsoft Corporation. (1993). Microsoft Excel (Version 5.0a) [Computer software]. Redmond, WA: Author.

Monroe, W. S. (1922). *Relation of sectioning a class to the effectiveness of instruction* (Bulletin No. 11). Urbana: University of Illinois, College of Education, Bureau of Educational Research.

*Moody, J. D., & Gifford, V. D. (1990, November). *The effect of grouping by formal reasoning ability, formal reasoning ability levels, group size, and gender on achievement in laboratory chemistry.* Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 326 443)

*Mortlock, R. S. (1969). Provision for individual differences in eleventh grade mathematics using flexible grouping based on achievement of behavioral objectives: An exploratory study (Doctoral dissertation, University of Michigan, 1969). *Dissertation Abstracts International, 30,* 3643A.

Mullen, B., & Copper, C. (1994). The relationship between group cohesiveness and performance: An integration. *Psychological Bulletin, 115,* 210–227.

Noddings, N. (1989). Theoretical and practical concerns about small groups in mathematics. *Elementary School Journal, 89,* 607–623.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

*Park, I. H. (1993). Cooperative learning and individual learning with computer-assisted instruction in an introductory university level chemistry course (Doctoral dissertation, University of Texas at Austin, 1993). *Dissertation Abstracts International, 54,* 2972A.

*Peterson, P. L., & Janicki, T. C. (1979). Individual characteristics and children's learning in large-group and small-group approaches. *Journal of Educational Psychology, 71,* 677–687.

*Peterson, P. L., Janicki, T. C., & Swing, S. R. (1981). Ability × treatment interaction effects on children's learning in large-group and small-group approaches. *American Educational Research Journal, 18,* 453–473.

*Petty, M. C. (1953a). *Intraclass grouping in the elementary school.* Unpublished doctoral dissertation, University of Texas, Bureau of Laboratory Schools.

*Petty, M. C. (1953b). *Intraclass grouping in the elementary school.* Austin: University of Texas Press.

Piaget, J. (1954). *The construction of reality in the child.* New York: Ballatine Books.

*Putbrese, L. M. (1972). An investigation into the effects of selected patterns of grouping upon arithmetic achievement (Doctoral dissertation, University of South Dakota, 1971). *Dissertation Abstracts International, 32,* 5113A.

Rosenbaum, J. E. (1976). *Making inequality: The hidden curriculum of high school tracking.* New York: Wiley.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64,* 479–530.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research, 13,* 89–99.

*Sandby-Thomas, M. (1983). The organization of reading and pupil attainment. *Journal of Research in Reading, 6,* 29–40.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105,* 309–316.

Shields, J. M. (1927). Teaching reading through ability-grouping. *Journal of Educational Methods, 7,* 7-9.

*Slavin, R. E. (1978). Student teams and comparison among equals: Effects on academic performance and student attitudes. *Journal of Educational Psychology, 70,* 532–538.

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research, 57,* 293–336.

Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research, 60,* 471–499.

*Slavin, R. E., & Karweit, N. L. (1984a). Mastery learning and student teams: A factorial experiment in urban general mathematics classes. *American Educational Research Journal, 21,* 725–736.

Slavin, R. E., & Karweit, N. L. (1984b, April). *Within-class ability grouping and student achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Slavin, R. E., & Karweit, N. L. (1985). Effects of whole class, ability grouped, and individualized instruction on mathematics achievement. *American Educational Research Journal, 22,* 351–367.

*Smith, W. M. (1961). The effect of intra-class ability grouping on arithmetic achievement in grades two through five (Doctoral dissertation, Louisiana State University, 1960). *Dissertation Abstracts International, 21,* 563.

Spence, E. S. (1958). Intra-class grouping of pupils for instruction in arithmetic in the intermediate grades of the elementary school (Doctoral dissertation, University of Pittsburgh, 1958). *Dissertation Abstracts International, 58,* 5635.

SPSS, Inc. (1994). SPSS 6.1 for Windows update [Computer software]. Chicago: Author.

*Stern, A. M. (1972). Intraclass grouping of low achievers in mathematics in the third and fourth grades (Doctoral dissertation, University of California, 1971). *Dissertation Abstracts International, 32,* 5539A.

*Terwel, J., Herfs, P. G. P., Mertens, E. H. M., & Perrenet, J. C. (1994). Cooperative learning and adaptive instruction in a mathematics curriculum. *Journal of Curriculum Studies, 26,* 217–233.

*Tingle, J. B., & Good, R. G. (1990). Effects of cooperative grouping on stoichiometric problem solving in high school chemistry. *Journal of Research in Science Teaching, 27,* 671–683.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes.* Cambridge, MA: Harvard University Press.

*Wallen, N. E., & Vowles, R. O. (1960). The effect of intraclass ability grouping on arithmetic achievement in the sixth grade. *Journal of Educational Psychology, 51,* 159–163.

*Watson, S. B. (1988). *Cooperative learning and group educational modules: Effects on cognitive achievement of high school biology students.* Unpublished doctoral dissertation, University of South Florida, Tampa, FL.

*Webb, N. M. (1982a). Group composition, group interaction, and achievement in cooperative small groups. *Journal of Educational Psychology, 74,* 475–484.

*Webb, N. M. (1982b). Peer interaction and learning in cooperative small groups. *Journal of Educational Psychology, 74,* 642–655.

*Webb, N. M. (1984). Stability of small group interaction and achievement over time. *Journal of Educational Psychology, 76,* 211–224.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13,* 21–39.

*Webb, N. M., Qi, S., Yan, K. X., Bushey, B., & Farivar, S. (1990, April). *Cooperative small-group problem solving in middle school mathematics.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Wilkinson, L. C. (1986, June). *Grouping low-achieving students for instruction.* Paper presented at the Designs for Compensatory Education Conference, Washington, DC. (ERIC Document Reproduction Service No. ED 293 915)

*Wright, S., & Cowen, E. L. (1985). The effects of peer teaching on student perceptions of class environment, adjustment, and academic performance. *American Journal of Community Psychology, 13,* 417–431.

*Yager, S., Johnson, D. W., & Johnson, R. (1985). Oral discussion, group-to-individual transfer, and achievement in cooperative learning groups. *Journal of Educational Psychology, 77,* 60–66.

*Yueh, J. S., & Alessi, S. M. (1988). The effect of reward structure and group ability composition on cooperative computer assisted instruction. *Journal of Computer Based Instruction, 15,* 18–22.

*Ziegler, S. (1981). The effectiveness of cooperative learning teams for increasing cross-ethnic friendship: Additional evidence. *Human Organization, 40,* 264–268.

*Zisk, J. F. (1993). The effects of a cooperative learning program on the academic self-concept of high school chemistry students (Doctoral dissertation, Temple University, 1993). *Dissertation Abstracts International, 54,* 3711A.

Authors

YIPING LOU is a PhD candidate, Centre for the Study of Classroom Processes, Concordia University, LB-581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; ylou@alcor.concordia.ca. She specializes in educational technology and research synthesis.

PHILIP C. ABRAMI is Professor and Director, Centre for the Study of Classroom Processes, Concordia University, LB-581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; pabrami@alcor.concordia.ca. He specializes in social psychology of education, research synthesis, and postsecondary instruction.

JOHN C. SPENCE is Research Associate, Alberta Centre for Well-Being, Faculty of Physical Education and Recreation, University of Alberta, Edmonton, Alberta, Canada T6G 2H9; j.spence@ualberta.ca. He specializes in exercise and sport psychology and quantitative synthesis.

CATHERINE POULSEN is Research Associate, Centre for the Study of Classroom Processes, Concordia University, LB-581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; cpoul@alcor.concordia.ca. She specializes in cooperative learning and second-language reading.

BETTE CHAMBERS is Associate Professor, Centre for the Study of Classroom Processes, Concordia University, LB-581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; bchamb@alcor.concordia.ca. She specializes in early childhood education and cooperative learning.

SYLVIA D'APOLLONIA is Professor, Centre for the Study of Classroom Processes, Concordia University, LB-581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; sdapoll@vax2.concordia.ca. She specializes in science education, meta-analysis, and reasoning.

Received January 2, 1996
Revision received June 20, 1996
Accepted July 26, 1996