

Meta-analytic Findings on Grouping Programs

James A. Kulik and Chen-Lin C. Kulik
The University of Michigan

Abstract

Meta-analytic reviews have focused on five distinct instructional programs that separate students by ability: multilevel classes, cross-grade programs, within-class grouping, enriched classes for the gifted and talented, and accelerated classes. The reviews show that effects are a function of program type. Multilevel classes, which entail only minor adjustment of course content for ability groups, usually have little or no effect on student achievement. Programs that entail more substantial adjustment of curriculum to ability, such as cross-grade and within-class programs, produce clear positive effects. Programs of enrichment and acceleration, which usually involve the greatest amount of curricular adjustment, have the largest effects on student learning. These results do not support recent claims that no one benefits from grouping or that students in the lower groups are harmed academically and emotionally by grouping.

Research on ability grouping has a long history. It goes back at least 75 years to 1916 when a researcher in Urbana, Illinois, studied the effects of special class placement on a group of high-aptitude 5th and 6th graders (Whipple, 1919). In the years since, researchers have carried out hundreds of additional studies of grouping, and reviewers have written dozens of reviews on the topic. Few educational practices have been scrutinized by researchers and reviewers for a longer period of time.

Despite all the effort, however, few clear cut conclusions have emerged. Some reviewers have concluded that research supports the practice of grouping. Others have concluded the opposite. And some reviewers have simply reported that clear-cut conclusions are impossible when research findings are so variable and conflicting. Basic questions therefore continue to trouble educators. Does anyone benefit from grouping? Who benefits most? Is anyone harmed? How? Why?

It seems to us that there are at least two good reasons for asking once again what the research says about such matters. First, social scientists have developed objective, scientific methods for research reviewing during the past decade (e.g., Glass, McGaw, & Smith, 1981). These methods were not available during the heyday of grouping research, and most older reviews therefore relied on impressionistic and subjective methods for summarizing and interpreting research literature. With the new scientific methods now available, we are finally in a good position to determine what the research actually says.

There is another reason for turning our attention once again to research on ability grouping. Ability grouping has come under strong attack recently. Oakes (1985), a leader in the

attack, has charged that ability grouping is discriminatory, unfair, and ineffective. In her view, no children gain from placement in homogeneous classes, and children in slower groups are harmed both intellectually and psychologically by grouping. Educators need to know whether the research actually supports such charges.

Our purpose therefore is to reexamine the findings on grouping using state-of-the-art methods for summarizing and interpreting the literature. We will examine findings on five distinct programs that separate students by ability:

1. *Multilevel classes.* Students in the same grade are divided into groups—often high, middle, and low groups—on the basis of ability, and the groups are instructed in separate classrooms either for a full day or for a single subject.
2. *Cross-grade grouping.* Children from several grades are formed into groups on the basis of their level of achievement in a subject, and the groups are then taught the subject in separate classrooms without regard to the children's regular grade placement.
3. *Within-class grouping.* A teacher forms ability groups within a single classroom and provides each group with instruction appropriate to its level of aptitude.

Putting the Research to Use

Advocates of *de-tracking* are today calling on schools to eliminate all forms of ability grouping. Meta-analytic results suggest that this proposed reform would damage American education. Teachers, counselors, administrators, and parents should be aware that student achievement would suffer from the wholesale elimination of school programs that group students by aptitude.

The harm would be relatively small from the simple elimination of multilevel classes, in which high, middle, and low groups cover the same curriculum. If schools replaced all their multilevel classes with mixed-ability ones, the achievement level of higher aptitude students would fall slightly, but the achievement level of other students would remain the same. If schools eliminated grouping programs with differentiated curricula, the damage to student achievement would be greater, and it would be felt broadly. Both higher and lower aptitude students would suffer academically from elimination of such programs. The damage would be truly great if, in the name of *de-tracking*, schools eliminated enriched and accelerated classes for their brightest learners. The achievement level of such students would fall dramatically if they were required to move at the common pace. No one can be certain that there would be a way to repair the harm that would be done.

4. *Enriched classes for the gifted and talented.* Students who are high in aptitude receive richer, more varied educational experiences than would be available to them in the regular curriculum for their age level.
5. *Accelerated classes for the gifted and talented.* Students who are high in academic aptitude receive instruction that allows them to proceed more rapidly through their schooling or to finish schooling at an earlier age than other students.

Our conclusion is that effects of grouping are a function of program type. Multilevel classes, which usually entail only minor adjustment of course content for ability groups, typically have little or no effect on student achievement. Programs that entail more substantial curricular adjustment, such as cross-grade and within-class programs, produce clear positive effects. Programs of enrichment and acceleration, which involve the greatest degree of curricular adjustment, have the largest effects on student learning. These results do not support recent claims that no one benefits from grouping and that students in the lower groups are harmed academically and emotionally by grouping.

Meta-analytic Methods

The review method used in this article is called *meta-analysis*. The method was first described in 1976 by Gene V. Glass in his presidential address to the American Educational Research Association. Meta-analysis is simply the analysis of analyses or, more formally, the application of quantitative statistics to the collected results of a large number of independent studies for the purpose of integrating the findings. To carry out a meta-analysis, a reviewer usually (a) finds as many studies as possible of an issue through an objective search of the literature; (b) codes the characteristics of these studies; (c) expresses the results of each study on a common metric; and (d) uses statistical methods to describe relationships between study characteristics and outcomes.

For each study included in a meta-analysis, the treatment effect is expressed in standard deviation units, or as an *effect size*. In principle, the computation of effect sizes is simple. A reviewer simply divides the gain or loss for an experimental group by an estimate of the population standard deviation on the outcome measure. An effect size is positive when there is a gain from the treatment and negative when there is a loss. An effect size is large when its absolute value is around 0.8, medium when around 0.5, and small when around 0.2.

We first used meta-analytic methods in 1982 to integrate research findings on ability grouping in secondary schools (Kulik & Kulik, 1982). We later extended our reviews to cover grouping in elementary schools (C. Kulik & Kulik, 1984), programs of accelerated instruction (J. Kulik & Kulik, 1984), and within-class and cross-grade grouping programs (Kulik & Kulik, 1987). Our most recent reports have provided an overview of this earlier work (e.g., Kulik & Kulik, 1991). Slavin (1987, 1990) applied his own variant of meta-analysis, called best-evidence synthesis, to both elementary school and secondary school findings on grouping.

Meta-analytic Results

The results reported in this article come from a recent updated statistical analysis that takes into account earlier meta-analytic work by both us and Slavin. The pool of studies used in the analysis is very similar to the combined pool of studies used in the two earlier sets of meta-analyses. It is not identical, however. We reread all the studies used in earlier analyses and reviewed Slavin's critique of various studies, and on this basis we eliminated from this new analysis a few studies included in earlier analyses. We also reviewed coding of all studies, and we revised our earlier coding when it seemed appropriate to do so. Our goal was to base conclusions in this report on the best interpretation of the best and most complete set of studies that we could assemble. A full report on our updated analysis of grouping findings is available from the National Research Center on the Gifted and Talented (Kulik, 1991).

Multilevel Classes

In 1919 Detroit became the first large city to introduce a formal multilevel plan of ability grouping (Courtis, 1925). The Detroit plan called for intelligence testing of all school children at the start of Grade 1 and then placement of children into X, Y, and Z groups on the basis of test results. The top 20% went to the X classes, the middle 60% to Y classes, and the bottom 20% to Z classes. Standard materials and methods were used in all classes, and no real adjustment of curriculum and methods was made for the ability groups.

Although many school systems followed the Detroit model and instituted three-tier grouping in subsequent years, their plans sometimes differed from the Detroit plan in significant ways. Few schools relied so exclusively on intelligence tests for initial placement in groups, and few separated students at such an early age. In addition, in many programs, especially those in high schools, the separation was not for a full day but was restricted instead to a single subject. Like the Detroit plan, however, most programs were set up simply to make things easier for teachers by reducing pupil variation in their classes. Few programs used multilevel classes as a way of providing differentiated curricula to the ability groups.

A total of 56 studies examined effects on students of placement in multilevel classes. A total of 51 of the 56 studies measured effects on achievement tests. Nearly 60% of the studies found higher examination scores in the multilevel classes; about 40% found higher examination scores in the mixed-ability classes. The difference in scores from homogeneous and mixed-ability classes was trivial or small, however, in virtually every one of the 51 studies. The average effect size in all programs was 0.03. This effect is only slightly less than the one (0.06) that we found in our earlier meta-analyses, and it is consistent with the effect of zero found by Slavin for multilevel grouping programs (Kulik & Kulik, 1991; Slavin, 1987, 1990). The effect is not large enough to be considered statistically different from zero.

A total of 36 of the 51 studies examined results separately by ability level. Effects varied slightly with aptitude. The average effect size was 0.10 for higher aptitude, -0.02 for middle aptitude,

and -0.01 for lower aptitude students. The average effect size of 0.10 for higher aptitude students was significantly greater than zero, and it was significantly higher than the average effect sizes for middle and lower aptitude students.

Effects showed little relation to study features. Results were very similar, for example, in true experiments and quasi-experimental studies, in studies reported in journals and dissertations, in studies evaluating full-day grouping programs and studies evaluating single-subject grouping, and so forth.

Thirteen of the 56 studies described effects of grouping on student self-esteem. The average overall effect of grouping in the 13 studies was to decrease self-esteem scores by 0.03 standard deviations, a very small and statistically nonsignificant amount. Eleven of the 13 studies also reported results separately by ability level. The average effect size was 0.19 for lower aptitude students, -0.09 for middle aptitude students, and -0.15 for higher aptitude students. Instruction in homogeneous classes thus tended to raise the self-esteem scores of lower aptitude students and to reduce the self-esteem of higher aptitude students.

Cross-grade grouping

The best known plan for cross-grade grouping is the Joplin plan. This grouping approach was devised by Cecil Floyd, who was then assistant superintendent of schools in Joplin, Missouri, and it was first used in the Joplin schools in 1953. The plan called for cross-grade grouping of fourth, fifth, and sixth graders for reading instruction. During the hour reserved for reading, children in these grades would break up into groups that went to reading classes on anything from the second- to the ninth-grade level. In these classes, the children would work with other fourth, fifth, and sixth graders who were reading at the same level. After this period was over, the children returned to their age-graded homerooms for a 25-minute period of reading for enjoyment. Almost all formal evaluations of cross-grade grouping involve the Joplin plan for reading instruction in elementary schools.

Cross-grade grouping is like multilevel grouping in that students of different ability levels are taught in separate classrooms. But in cross-grade plans, there are typically more levels. In a typical Joplin program, for example, a fifth grader might be assigned to any one of nine different reading groups. In addition, cross-grade grouping is single-subject grouping, and so group placement is usually tied closely to a specific skill. Perhaps the most important difference between cross-grade and multilevel grouping, however, is in the amount of curricular adjustment in the two approaches. In cross-grade programs, students in different ability groups work with different materials and different methods. In most multilevel programs, little or no effort is made to adjust curriculum to group ability level.

Fourteen studies investigated effects of such cross-grade programs. Eleven of the studies found that students achieved more when taught in these cross-grade programs; two studies found that performance was better when students were taught in conventional mixed-ability classes; and one study found no difference in results of the two approaches. The average effect

size in the 14 studies was 0.30 , a small effect but one that is significantly greater than zero.

Two of the studies reported results separately by ability level. The average effect was 0.12 for the high-ability students; -0.01 for the middle-ability students; and 0.29 for the low-ability students. Because of the small number of studies of cross-grade grouping in the literature, it was not possible to examine further the relationship between study characteristics and study outcomes. None of the studies reported on effects of cross-grade grouping on self-esteem.

Within-class grouping

Elementary school teachers often group the children in a class into subgroups for specific activities and purposes. They use such subgroups especially often for reading and arithmetic lessons, and they sometimes form subgroups for science and social science projects as well. The teacher usually presents a lesson to one of the subgroups while the remaining groups engage in other activities.

Two facts about within-class grouping plans make them especially interesting. First, most within-class grouping plans call for differentiated instruction for the groups. For the practice of within-class grouping to make sense, the teacher must present different material to each group. It would be inefficient for a teacher to divide a class into thirds on the basis of ability and then to make the same presentation separately to each of the three groups. Thus, within-class programs are like cross-grade programs in that they involve differentiated curriculum. Second, within-class programs do not involve assignment of groups to separate classrooms. Within-class programs differ from both multilevel and cross-grade programs in this respect.

Eleven studies described results from within-class grouping programs. Nine of these studies reported a higher overall achievement level with within-class grouping; only two studies reported a higher overall achievement level with mixed-ability instruction. The average overall effect of grouping in the 11 studies was to raise examination scores by 0.25 standard deviations, a significant but small effect.

Six of the 11 studies reported results separately by ability group. Effects were small to moderate for students at all ability levels. The average effect size was 0.30 for the higher ability students; 0.18 for the middle ability students; and 0.16 for the low-ability students. Too few studies were available for analysis of the relationship between study features and effect sizes.

Enriched Classes for the Gifted and Talented

Some grouping programs are designed especially to meet the needs of gifted and talented students. Learners in these programs are ordinarily a distinctive group with unusually high academic aptitude. Teachers in such programs usually believe that their students have special needs, and they usually have a strong commitment to meeting these needs. The result is typically a highly challenging educational program with distinctive materials and methods adapted to student ability.

We found a total of 25 studies of special programs for the gifted and talented. Twenty-two of the 25 studies found that

talented students achieved more when they were taught in special programs. The average effect in the 25 studies was 0.41. This effect is moderate in size and significantly greater than an effect size of zero. We were unable to find any study feature that was significantly related to variation in effect size. The small number of studies available for analysis might account in part for this failure to find significant relationships.

Five of the 25 studies of special programs for the gifted and talented investigated effects on self-concept. In 4 of the 5 studies, self-concepts were more favorable when the gifted and talented were taught in separate groups. The size of the effect was small or trivial, however, in all of the studies. The average effect size in all 5 studies was 0.10.

Accelerated Classes for the Gifted and Talented

Acceleration of the gifted can take a variety of forms. Some programs entail radical acceleration of individual students; some involve more moderate advancement of groups of students. We were able to find 23 controlled studies of accelerated instruction in the literature. The 23 studies did not examine effects of radical acceleration but instead evaluated more modest forms of rapid advancement. These included compressing a curriculum for talented students (e.g., 4 years in 3) and extending the calendar to speed up the progress of such students (e.g., completing the work of 4 years in 3 school years with five summer sessions).

The 23 studies used two different study designs that reflected fundamentally different research purposes. In one group of studies, the groups being compared were initially equivalent in age and aptitude, but because one group was accelerated and the other was not, the two groups differed in grade level when educational outcomes were measured. In a second group of studies, accelerated students were compared with older, highly talented nonaccelerates in the grades into which the accelerates had moved.

These two types of studies produced distinctly different results. Each of the 11 studies with same-age control groups showed greater achievement in the accelerated class; the average effect size in these studies was 0.87. Studies with older comparison groups were as likely to produce positive as negative differences between groups. The average effect size in the 12 studies with older comparison groups was -0.02.

Only a small number of studies investigated other outcomes of acceleration, and findings were not entirely consistent from study to study. On the average, however, acceleration appeared to have little or no effect on students' attitude toward school, participation in school activities, popularity, or adjustment. Acceleration had a strong effect on vocational plans in two studies but trivial effects on vocational plans in four other studies. The effect on vocational plans apparently varied as a function of program type.

Discussion and Conclusion

Oakes (1985) concluded that no one gains academically from ability grouping and that lower aptitude children lose a good deal of academic ground when taught in homogeneous groups. Our analyses do not support this conclusion. They point instead

to some clear and consistent academic benefits from grouping programs. The academic benefits are clearest for those in the higher ability groups, but students in the lower groups are not harmed academically by grouping and they gain academic ground in some grouping programs.

For all types of students, however, the size of academic gains is a function of program type. Multilevel classes generally have little or no effect on student achievement levels; within-class and cross-grade programs generally produce small positive effects; enriched and accelerated classes produce moderate-to-large positive effects. Several different factors might account for the difference in program results. We believe, however, that the key factor is the degree to which course content is adjusted to group ability in the programs.

Reports on multilevel classes seldom describe planned curricular adjustment to group ability. For some of the older studies of multilevel classes, in fact, teachers were told to keep content constant across ability groups. Even in more recent studies of multilevel classes, adjustment of content to ability group is usually informal and at the discretion of individual teachers. In contrast, reports on cross-grade and within-class programs usually describe planned adjustment of content to group ability. In cross-grade programs, students move up or down grades for reading instruction to ensure a match between their ability and their reading instruction. In within-class programs, teachers divide students into ability groups so that they can work on different materials with children of differing ability levels.

Enriched and accelerated classes are by definition classes in which material is adjusted to the needs of special groups. In enriched classes, the emphasis is on giving students a richer and more varied educational experience than they would receive in regular classes. In accelerated classes, the emphasis is on providing instruction that allows children to proceed more rapidly through schooling or to finish at an earlier age. Gains on tests are larger for children in accelerated classes, but gains in enriched classes also seem impressive when one considers the special emphases in the classes. In some enriched classes, children spend as much as half their time on cultural material (e.g., foreign languages, music, art) not covered on standard achievement tests.

It is also important to note that ability grouping does not have devastating effects on student self-esteem, as Oakes (1985) has charged. Effects of grouping on self-esteem are near-zero overall. They appear to be slightly positive for lower ability students and slightly negative for higher aptitude ones. Talented students may become slightly less satisfied with themselves when taught with their intellectual peers; slower students may gain slightly in self-confidence when they are taught with other slower learners.

Our conclusions are therefore very different from those reached by Oakes. Whereas Oakes concludes that grouping programs are unnecessary, ineffective, and unfair, we conclude that the opposite is true. We believe that American schools would be harmed by the elimination of programs that tailor instruction to the aptitude, achievement, and interests of groups with special educational needs.

References

Courtis, S. A. (1925). Ability-grouping in Detroit schools. In G. M. Whipple (Ed.), *The ability grouping of pupils*, 35th Yearbook of the National Society for the Study of Education (Part I, pp. 44-47). Bloomington, IL: Public School Publishing.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.

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

Kulik, C.-L. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, 19, 415-428.

Kulik, C.-L., & Kulik, J. A. (1984, August). *Effects of ability grouping on elementary school pupils: A meta-analysis*. Paper presented at the annual meeting of the American Psychological Association, Toronto. (ERIC Document Reproduction Service No. Ed 255 329)

Kulik, J. A. (1991). *Ability grouping*. Research-based decision-making series. Storrs, CT: National Research Center on the Gifted and Talented, University of Connecticut.

Kulik, J. A., & Kulik, C.-L. C. (1984). Effects of accelerated instruction on students. *Review of Educational Research*, 54, 409-426.

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). Ability grouping and gifted students. In N. Colangelo & G. Davis (Eds.), *Handbook of gifted education* (pp. 178-196). Boston, MA: Allyn & Bacon.

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

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: 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.

Whipple, G. M. (1919). *Classes for gifted children*. Bloomington, IL: Public School Publishing.

**CREATE-A-CENTER
ABOUT...SERIES**

Designed with the gifted student in mind, the challenging, multi-disciplinary activities encourage students' imaginations to soar!

Each center features 90 tasks on attractive cards plus bonus materials and sells for \$18.95.



A few of the titles include . . .

- America Becomes a Nation • Astronomy • Dinosaurs
- The Environment • The Future • Greek Mythology
- The Middle Ages • Native Americans • Primary Topics

Write or call for your free catalog describing these and other resources for teachers of the gifted and talented.

Educational Impressions

210 Sixth Avenue, P.O. Box 77, Hawthorne, NJ 07507
(201) 423-4666 (800) 451-7450 FAX (201) 423-5569

Quality Materials to Enrich the Curriculum and Encourage Critical and Creative Thinking!