Findings on Grouping Are Often Distorted
Response to Allan

I agree with Allan that the research on grouping is often misinterpreted—to the disservice of our students.

Susan Allan makes several important contributions to the controversy about research findings on grouping. She identifies questions that quantitative reviewers have and have not asked about the practice. She describes the review findings and distinguishes these actual findings from misinterpretations in the professional and popular literature. She also points to design weaknesses that afflict many studies of grouping. Arguments about grouping have become heated and loud in recent years. Allan’s calm voice provides some welcome relief.

One of her most important points is that “grouping” is not a single thing. Rather, it comes in a variety of forms and is done for a variety of reasons. It is, therefore, a mistake to think that the different approaches to grouping all have the same effect. In my view, reviewers and educators should distinguish among at least three types of programs:

Type I: simple programs in which all ability groups are taught with the same or similar materials and methods. Such programs usually produce small positive gains in student learning: student scores rise on the average about 0.1 of a standard deviation, or about one month on a grade-equivalent scale.

Type II: programs in which teaching materials and methods are adjusted to meet the special needs of a specific aptitude group (for example, enriched instruction for the gifted and talented). Type II programs produce larger effects: student scores rise on the average about 0.4 of a standard deviation, or about four months on a grade-equivalent scale.

Type III: programs in which adjustment of teaching materials is so extensive that it affects a student’s rate of progress through school (for example, programs of accelerated instruction). Type III programs produce the strongest effects of all: gains of one full standard deviation, or one year on a grade-equivalent scale, are not uncommon.

The criterion measures used in research studies of grouping, however, may not be entirely adequate; these figures may, therefore, underestimate the size of the true effects. As Allan points out, almost all studies use standardized tests as criterion measures, and such tests may not provide the best measures of achievement in specific school systems. Both Slavin (1987) and Kulik and Kulik (1991) have shown that results from studies using standardized tests tend to be weaker than those from studies using local tests as criterion examinations.

Allan’s major worry about grouping research, however, is not its methodological weakness but rather its misinterpretation. I share this concern. Most evaluations have focused on Type I programs. The evidence that these programs usually lead to small positive gains in student learning has been twisted, however, to support the conclusion that grouping programs do not work and thus should be eliminated. This blanket condemnation of grouping has been extended to Type II programs, even though the evidence on these programs is clearly favorable. Even Type III programs are in jeopardy in some school systems. Our children will be the losers if reviewers continue to twist research findings to fit their personal and political philosophies.

References


James A. Kulik is Research Scientist, Center for Research on Learning and Teaching, The University of Michigan, 109 E. Madison St., Ann Arbor, MI 48109.