

From the Editor—

One of the central issues of research is making an appropriate comparison. Results from a treatment are meaningless by themselves. To have meaning, they must be compared to something. Knowing that a student got 10 items correct on a test tells us nothing unless we have some context for comparison; that context could be a mastery criterion, the student's performance on a pretest, the performance of similar students, and so on. Each of these kinds of comparisons gives different kinds of meaning to the raw score.

When we evaluate the results from a study, we must always be aware of the comparisons that are being made. Research studies differ in the kinds of comparisons that are made and therefore, the kinds of conclusions that are warranted. This issue of the *Journal of Direct Instruction (JODI)* features articles with several different kinds of comparisons. Ginn, Keel, and Fredrick compare a treatment group to several control groups. All the students were enrolled in a program for gifted and talented students; those at one school used *Reasoning and Writing* and those at three other schools used more typical gifted and talented methods and curriculum. The use of three separate control schools is particularly interesting because it reveals that the gains made by students in the *Reasoning and Writing* class were outside the range of typical variability across schools. When sufficient data are available (for example when using district-wide or statewide tests), this method could be used more broadly. The use of several comparison groups contributes an additional dimension to

the comparisons that we can make, and the conclusions that we can draw.

Unfortunately, relevant control groups are not always available. If we use tests that are not already administered to potential comparison groups, it can be very difficult to convince these groups to participate in testing. (Ginn, Keel, and Fredrick are to be congratulated for overcoming this hurdle.) When control groups are not available, the next best alternative may be to compare participants' growth from pretest to posttest to the amount of growth that would be expected based on the test's norms. The problem with this approach is that there may be important differences between the treatment group and the group on which the test was normed. For this reason, we must be more cautious in drawing conclusions from this kind of study. Any single study that uses test norms as a comparison can support only tentative conclusions. But an accumulation of studies that show similar results can justify stronger conclusions. Three studies in this issue of *JODI* made comparisons to test norms because the relevant tests were not given more broadly. Anderson and Keel evaluate the effects of *Reasoning and Writing* on students with learning disabilities by comparing students' relative standing compared to the norm group on a pretest to their relative standing on a posttest. Similarly, Fredrick, Keel, and Neel examine the progress of students in *Reading Mastery* by comparing students' pretest standing to their posttest standing. Scarlato and Burr investigate the results of using teacher-designed lessons based on Direct Instruction principles for teaching mathematics. This study makes pretest and posttest comparisons to the norms of a standardized test and also uses a very small comparison group.

In a very different kind of study, Cross, Rebarber, and Wilson also make use of test norms as a comparison. This study presents a large-scale evaluation of achievement across a school year for students who attended schools administered by a private company— Advantage Schools. This evaluation involves

thousands of students and uses well-known standardized tests. The problem with this kind of evaluation is that a relevant control group is difficult to identify, much less to assess. Thus, the authors compared progress of students in Advantage Schools to the progress that would be expected based on the tests' norms.