Response to Greenleaf and Petrosino

I am pleased to respond to concerns raised by Cynthia Greenleaf and Anthony Petrosino (2009) in their letter to the editors about our review of research on reading programs for middle and high schools. One of our hopes in writing the review was to open up just such conversations about methods for reviewing "what works" in the teaching of reading.

One of the issues raised by Greenleaf and Petrosino relates to our categorization of programs as reading curricula, CAI, instructional process, or mixed methods. They are correct in saying that sometimes programs are difficult to categorize, and there is no mechanistic way to resolve close calls. However, most decisions were not difficult. For example, Greenleaf and Petrosino sought clarification on our distinction between CAI (such as *Jostens*) and mixed-methods (such as *Read 180*). Programs in both categories do use computers, but otherwise they are very different, as the CAI programs are invariably supplementary, are used less than two hours a week, and involve modest professional development. *Read 180* and *Voyager Passport* use many non-computer group activities as well as computer activities, provide teachers with significant professional development and follow-up, and are intended to serve as core, not supplementary, instruction for struggling readers. In the case of cooperative learning and metacognitive strategies, it is true that all of the cooperative learning programs also focused on metacognitive strategies, but not all of the metacognitive strategy programs emphasized cooperative learning. Perhaps we should have called the cooperative learning methods "cooperative learning and metacognitive strategies," but I think the distinction was clear.

Greenleaf and Petrosino raise questions about pooling of effect sizes from multiple grade levels and subgroups. Our procedures for pooling were consistent. We averaged effect sizes across measures within grades and then across grades. We did average SAT-9 and Terra Nova outcomes from the three cohorts of the White, Haslam, and Hewes (2006) and Johnson, Haslam, and White (2006) study, for example. We did not average in follow-up measures, however, which is why the AIMS scores are reported but not included in the mean. The reason for this is that because follow-up measures usually show diminished effect sizes, averaging them into the main summary effect would unfairly penalize programs that report follow-up measures. When there were multiple years of intervention for the same students, we used data from the final year, so in the Gaskins (1994) evaluation of *Benchmark Detectives* we used only the data from the second-year assessments.

The idea that there may be different expected outcomes at different grade levels may be valid, but it is impossible to deal with as a reviewer. It may be more difficult to raise reading scores in high school than in middle school, but reviewers still have to report whatever the effect sizes were. When there are more senior high school studies it would be interesting to explore this question further.

I agree with Greenleaf and Petrosino that subgroup effects are crucial, but as they also acknowledge, there are inadequate numbers of studies of particular subgroups to permit firm conclusions about differential effects. We show the subgroup effects when they were reported in the original articles and characterize the nature of the subjects, so it is possible to derive clues for future research to investigate.

I agree that pooling across outcome measures is less than ideal. In a related review of beginning reading programs (Slavin et al., 2009), we broke down the outcomes by decoding, vocabulary, and comprehension/total reading. We did not do this in the secondary review, but the information on different outcomes is reported, so readers are able to focus on particular types of outcomes if they so choose. I completely agree with Greenleaf and Petrosino about the importance of context and of the nature of the control groups. We have briefly described both of these issues, but space limitations make further detail difficult. Future reviewers may find sufficient numbers of studies that vary in context and in the nature of control groups, but at present it is difficult to form useful generalizations. Readers can of course go to the original studies to learn more about these important questions.

I wholly endorse the general conclusions made by Greenleaf and Petrosino, which I would paraphrase as (a) we need many more high-quality experimental studies of secondary reading programs, but (b) what we do already know gives us important tools in helping struggling secondary readers. Nothing would make me happier than to have future reviewers add information on the conditions under which various programs do and do not work, on the subgroups most and least helped by various programs, and contextual information on the entire process of reform. My colleagues and I have done our best to draw from the

limited evidence base some principles of best practice, but future work may well find that we were wrong on some or perhaps most of our conclusions. However, we believe that it is useful to try to characterize what we do know at this early stage as a service to educators who have to make important decisions today and hopefully as a spur to researchers and funders to add critical knowledge that we still lack.

Robert E. Slavin, Johns Hopkins University, Baltimore, MD, USA; University of York, England.

References

Gaskins, I.W. (1994). Classroom applications of cognitive science: Teaching poor readers how to learn, think, and problem solve. In K. McGilly (Ed.), *Classroom lessons* (pp. 129–154). Cambridge, MA: MIT Press.

Greenleaf, C., & Petrosino, A. (2009). Letters to the editors: Response to Slavin, Cheung, Groff, and Lake (2008). Effective Reading Programs for Middle and High Schools: A Best-Evidence Synthesis. *Reading Research Quarterly, 44*(4), 349–354.

Johnson, J., Haslam, M., & White, R. (2006). *Improving student literacy in the Phoenix Union High School District, 2005–06.* Washington, DC: Policy Studies Associates.

Slavin, R.E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). *Effective beginning reading programs: A best-evidence synthesis.* Baltimore, MD: Center for Research and Reform in Education, Johns Hopkins University.

White, R.N., Haslam, M.B., & Hewes, G.M. (2006, July). *Improving student literacy in the Phoenix Union High School District 2003–04 and 2004–05: Final Report.* Washington, DC: Policy Studies Associates.