Is There Better Evidence on Mastery Learning?
A Response to Slavin

James A. Kulik,
Chen-Lin C. Kulik,
University of Michigan

and

Robert L. Bangert-Drowns
State University of New York

Our purpose in writing “Effectiveness of Mastery Learning Programs: A Meta-Analysis” was to present as full a picture as possible of evaluation findings on mastery learning (Kulik, Kulik, & Bangert-Drowns, 1990). We wanted to get out all the facts on mastery learning so that we and others could draw conclusions about its effectiveness. We therefore described a variety of results from a variety of programs. For example, we covered cognitive and affective outcomes; effects of programs on course completion and student study time; findings from Grade 1 through college; and results from programs modeled after Bloom’s system of Learning for Mastery (LFM) and Keller’s Personalized System of Instruction (PSI). Further, we examined effects on average performance and variation in performance. And we looked at main effects and interactions.

If we can judge from his published writing, Slavin’s interest in mastery learning is narrower than ours (Slavin, 1987, 1990). He focuses exclusively on examination effects from precollege studies of Bloom’s LFM. He thus ignores evidence on college applications of LFM, although more than half of all evaluations of LFM are carried out in colleges. He also ignores Keller’s PSI, although two-thirds of the evaluations of mastery teaching involve PSI. Slavin does not look at affective outcomes of mastery teaching, nor does he look at its effects on variation in achievement. He gives scant attention to aptitude-achievement interactions and measures of instructional time. Thus our estimate is that Slavin’s focus is on no more than 10 percent of the mastery picture.

Although we believe that researchers need to see the whole picture to answer questions about mastery, we are restricting our comments here to the 10 percent of the story that interests Slavin: the effects of Bloom’s LFM on precollege examination scores. We first try to determine whether Slavin agrees with us on the facts about studies of precollege LFM. Our conclusion is that he does not dispute the facts that we presented in our review. We then attempt to formulate some
conclusions that are consistent both with these facts and with what we and Slavin value in research studies.

Two facts loom large in the evaluation literature on precollege LFM. First, LFM raises student scores on locally developed criterion examinations by moderate amounts. Second, LFM has much smaller effects on scores on standardized examinations. These are points about precollege LFM that we emphasized in our review, and they are also the points that Slavin stresses in his reply.

Slavin does not dispute some other important facts brought out in our review, but he does not face all of the implications of these facts. As our review pointed out, Slavin made some odd choices in selecting studies for his best-evidence synthesis. First, in several studies that he selected as exemplary evaluations of LFM, neither experimental nor control students were taught for mastery (e.g., Gutkin, 1985; Wyckoff, 1974). Second, in several other studies that he chose as exemplary, both experimental and control groups were mastery groups (e.g., Dunkelberger & Heikkinen, 1984; Fuchs, Tindal, & Fuchs, 1985). The first type of study lacks a mastery treatment group; the second type lacks a nonmastery control. Size of mastery effects cannot be estimated from either type of study, and Slavin should not have used these studies for this purpose. A review that estimates size of mastery effects from such irrelevant studies does not merit the epithet “best-evidence.”

Although Slavin agrees with us on the facts, he disagrees about which facts should be emphasized in drawing overall conclusions about Bloom’s LFM. Slavin thinks that our conclusions are misleading because they are based in part on results from three- and four-week studies and studies with only one experimental and one control group. He also thinks that we give too little emphasis to results of studies with standardized tests. We believe, on the other hand, that Slavin’s conclusions are distorted because they are based in part on results from studies with either no mastery treatment or no control treatment. We also think that Slavin gives too much emphasis to results from standardized tests.

But we do not consider our differences to be irreconcilable. We think it is possible to resolve them by (a) restricting discussion to the 11 studies that seem adequate to both sides and (b) examining separately effects on local and standardized tests in these studies. Table 1 presents estimates of effect sizes in the 11 studies. Median and mean effects are very similar for the estimates in the table. When Slavin’s calculations of effect size are used, the average effect size for local tests is 0.36 ($t(8) = 2.94, p < .05$), and the average effect for standardized tests is 0.09 ($t(4) = 2.30, p < .10$). When our calculations are used, the average effect size for local tests is 0.45 ($t(8) = 3.49, p < .01$), and the average effect for standardized tests is 0.08 ($t(4) = 3.00, p < .05$). It is notable that LFM produces statistically significant effects on both local and standardized tests, but the effects on local tests are obviously larger and more important. All things considered, the average effect of precollege LFM appears to be an increase of about 0.4 standard deviations on local tests and 0.1 standard deviations on standardized tests.

Too few studies are available, however, for us to feel confident about these estimates of LFM effect sizes. We would therefore like to see more research on LFM effects, especially its effects on standardized tests. Until more studies are conducted, researchers should keep in mind the current results: LFM students clearly do better than other students on tests developed to fit local curricula, and
Better Evidence on Mastery Learning

TABLE 1

Effect sizes in 11 studies of LFM at the precollege level

<table>
<thead>
<tr>
<th>Study</th>
<th>ES on local test</th>
<th>ES on standardized test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kulik et al.</td>
<td>Slavin</td>
</tr>
<tr>
<td>Anderson, Scott, &amp; Hutlock, 1976</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>Cabezón, 1984</td>
<td>0.84</td>
<td>0.27</td>
</tr>
<tr>
<td>Fagan, 1976</td>
<td>-0.11</td>
<td>-0.11</td>
</tr>
<tr>
<td>Jones, Monsaas, &amp; Katims, 1979</td>
<td>0.41</td>
<td>0.18</td>
</tr>
<tr>
<td>Katims, Smith, Steele, &amp; Wick, 1977</td>
<td>0.17</td>
<td>0.25</td>
</tr>
<tr>
<td>Kersh, 1970</td>
<td>-0.19</td>
<td>(-0.19)</td>
</tr>
<tr>
<td>Long, Okey, &amp; Yeany, 1978</td>
<td>0.43</td>
<td>0.43</td>
</tr>
<tr>
<td>Lueckmeyer &amp; Chiappetta, 1981</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>Mevarech, 1980</td>
<td>0.79</td>
<td>0.77</td>
</tr>
<tr>
<td>Mevarech, 1986</td>
<td>0.87</td>
<td>0.90</td>
</tr>
<tr>
<td>Slavin &amp; Karweit, 1984</td>
<td></td>
<td>0.02</td>
</tr>
</tbody>
</table>

Note: Values in parentheses have been plugged because Slavin noted only direction of difference.

They do only slightly better than others on standardized tests that sample objectives from many school systems and many grade levels. There is no evidence to suggest that LFM has any negative effects on any type of student learning.

Finally, teachers and researchers can learn much more about mastery learning by looking at the full picture rather than the small corner of the research on which Slavin focuses. They can find more stable and reliable estimates of effect sizes in the larger literature on mastery learning. And they can learn about affective and behavioral outcomes of mastery learning; about the conditions under which mastery teaching has stronger and weaker effects; about the students in a classroom who are most affected by a switch to mastery methods; and so on. We believe LFM effects on precollege examinations constitute only a small part of the story.

References


Kulik, Kulik, and Bangert-Drowns


Authors

JAMES A. KULIK is Research Scientist, Center for Research on Learning and Teaching, 109 E. Madison St., Ann Arbor, MI 48109. He specializes in research synthesis and educational evaluation.

CHEN-LIN C. KULIK is Assistant Research Scientist, Center for Research on Learning and Teaching, 109 E. Madison St., Ann Arbor, MI 48109. She specializes in research synthesis and psychological measurement.
Better Evidence on Mastery Learning

ROBERT L. BANGERT-DROWNS is Assistant Professor, University of Albany/SUNY, ED 313, 1400 Washington Ave., Albany, NY 12222. He specializes in critical thinking, instructional design, and research synthesis.