Review of Educational Research
Summer 1987, Vol. 57, No. 2, pp. 175–213

Mastery Learning Reconsidered

Robert E. Slavin
Johns Hopkins University

Several recent reviews and meta-analyses have claimed extraordinarily positive effects of mastery learning on student achievement, and Bloom (1984a, 1984b) has hypothesized that mastery-based treatments will soon be able to produce "2-sigma" (i.e., 2 standard deviation) increases in achievement. This article examines the literature on achievement effects of practical applications of group-based mastery learning in elementary and secondary schools over periods of at least 4 weeks, using a review technique, "best-evidence synthesis," which combines features of meta-analytic and traditional narrative reviews. The review found essentially no evidence to support the effectiveness of group-based mastery learning on standardized achievement measures. On experimenter-made measures, effects were generally positive but moderate in magnitude, with little evidence that effects maintained over time. These results are discussed in light of the coverage versus mastery dilemma posed by group-based mastery learning.

The term "mastery learning" refers to a large and diverse category of instructional methods. The principal defining characteristic of mastery learning methods is the establishment of a criterion level of performance held to represent "mastery" of a given skill or concept, frequent assessment of student progress toward the mastery criterion, and provision of corrective instruction to enable students who do not initially meet the mastery criterion to do so on later parallel assessments (see Block & Anderson, 1975; Bloom, 1976). Bloom (1976) also includes an emphasis on appropriate use of such instructional variables as cues, participation, feedback, and reinforcement as elements of mastery learning, but these are not uniquely defining characteristics; rather, what defines mastery learning approaches is the organization of time and resources to ensure that most students are able to master instructional objectives.

There are three primary forms of mastery learning. One, called the Personalized System of Instruction (PSI) or the Keller Plan (Keller, 1968), is used primarily at the postsecondary level. In this form of mastery learning, unit objectives are established for a course of study and tests are developed for each. Students may take the test (or parallel forms of it) as many times as they wish until they achieve a passing score. To do this, students typically work on self-instructional materials and/or work with peers to learn the course content, and teachers may give lectures

---

This paper was written under a grant from the Office of Educational Research and Improvement, U.S. Department of Education (No. OERI-G-86-0006). However, any opinions expressed are those of the author and do not represent OERI policy.

I would like to thank John Waters and Suzanne Shanahan for their assistance in locating references, and Jere Brophy, Barak Rosenshine, Nancy Madden, Nancy Karweit, Daniel Levine, and John Hollifield for their comments on earlier drafts.

more to supplement than to guide the learning process (see Kulik, Kulik & Cohen, 1979). A related form of mastery learning is *continuous progress* (e.g., Cohen, 1977), where students work on individualized units entirely at their own rate. Continuous progress mastery learning programs differ from other individualized models only in that they establish mastery criteria for unit tests and provide corrective activities for students who do not meet these criteria the first time.

The third form of mastery learning is called *group-based mastery learning*, or Learning for Mastery (LFM) (Block & Anderson, 1975). This is by far the most commonly used form of mastery learning in elementary and secondary schools. In group-based mastery learning the teacher instructs the entire class at one pace. At the end of each unit of instruction a "formative test" is given, covering the unit's content. A mastery criterion, usually in the range of 80–90% correct, is established for this test. Any students who do not achieve the mastery criterion on the formative test receive corrective instruction, which may take the form of tutoring by the teacher or by students who did achieve at the criterion level, small group sessions in which teachers go over skills or concepts students missed, alternative activities or materials for students to complete independently, and so on. In describing this form of mastery learning, Block and Anderson recommend that corrective activities be different from the kinds of activities used in initial instruction. Following the corrective instruction, students take a parallel formative or "summative" test. In some cases only one cycle of formative test-corrective instruction-parallel test is used, and the class moves on even if several students still have not achieved the mastery criterion; in others, the cycle may be repeated two or more times until virtually all students have gotten a passing score. All students who achieve the mastery criterion at any point are generally given an "A" on the unit, regardless of how many tries it took for them to reach the criterion score.

The most recent full-scale review of research on mastery learning was published more than a decade ago by Block and Burns (1976). However, in recent years two meta-analyses of research in this area have appeared, one by Kulik, Kulik, and Bangert-Drowns (1986) and one by Guskey and Gates (1985, 1986). Meta-analyses characterize the impact of a treatment on a set of related outcomes using a common metric called "effect size," the posttest score for the experimental group minus that for the control group divided by the control group's standard deviation (see Glass, McGaw, & Smith, 1981). For example, an effect size of 1.0 would indicate that, on the average, an experimental group exceeded a control group by one standard deviation; the average member of the experimental group would score at the level of a student in the 84th percentile of the control group's distribution.

Both of the recent meta-analyses of research on mastery learning report extraordinary positive effects of this method on student achievement. Kulik et al. (1986) find mean effect sizes of 0.52 for pre-college studies and 0.54 for college studies. Guskey and Gates (1985) claim effect sizes of 0.94 at the elementary level (grades 1–8), 0.72 at the high school level, and 0.65 at the college level. Further, Walberg (1984) reports a mean effect size of 0.81 for "science mastery learning" and Lysakowski and Walberg (1982) estimate an effect size for "cues, participation, and corrective feedback," principal components of mastery learning, at 0.97. Bloom (1984b, p. 7) claims an effect size of 1.00 "when mastery learning procedures are done systematically and well" and has predicted that forms of mastery learning will be able to consistently produce achievement effects of "2 sigma" (i.e., effect sizes of 2.00). To put these effect sizes in perspective, consider that the mean effect size for randomized studies of one-to-one adult tutoring reported by Glass, Cohen, Smith, and Filby (1982) was 0.62 (see Slavin, 1984a). If the effects of mastery learning instruction approach or exceed those for one-to-one tutoring, then mastery learning is indeed a highly effective instructional method.

The purpose of the present article is to review the research on the effects of group-based mastery learning on the achievement of elementary and secondary students in an attempt to understand the validity and the practical implications of these findings. The review uses a method for synthesizing large literatures called "best-evidence synthesis" (Slavin, 1986a), which combines the use of effect size as a common metric of treatment effect with narrative review procedures. Before synthesizing the "best evidence" on practical applications of mastery learning, the following sections discuss the theory on which group-based mastery learning is based, how that theory is interpreted in practice, and problems inherent in research on the achievement effects of mastery learning.

Mastery Learning in Theory and Practice

The theory on which mastery learning is based is quite compelling. Particularly in such hierarchically organized subjects as mathematics, reading, and foreign languages, failure to learn prerequisite skills is likely to interfere with students' learning of later skills. For example, if a student fails to learn to subtract, he or she is sure to fail in learning long division. If instruction is directed toward ensuring that nearly all students learn each skill in a hierarchical sequence, then students will have the prerequisite skills necessary to enable them to learn the later skills. Rather than accepting the idea that differences in student aptitudes will lead to corresponding differences in student achievement, mastery learning theory holds that instructional time and resources should be used to bring all students up to an acceptable level of achievement. To put it another way, mastery learning theorists suggest that rather than holding instructional time constant and allowing achievement to vary (as in traditional instruction), *achievement level* should be held constant and *time* allowed to vary (see Bloom, 1968; Carroll, 1963).

In an extreme form, the central contentions of mastery learning theory are almost tautologically true. If we establish a reasonable set of learning objectives and demand that every student achieve them at a high level *regardless of how long that takes*, then it is virtually certain that all students will ultimately achieve that criterion. For example, imagine that students are learning to subtract two-digit numbers with renaming. A teacher might set a mastery criterion of 80% on a test of two-digit subtraction. After some period of instruction, the class is given a formative test, and let's say half of the class achieves at the 80% level. The teacher might then work with the "nonmasters" group for one or more periods, and then give a parallel test. Say that half of the remaining students (25% of the class) pass this time. If the teacher continues this cycle indefinitely, then all or almost all students will ultimately learn the skill, although it may take a long time for this to occur. Such a procedure would also accomplish two central goals of mastery learning, particularly as explicated by Bloom (1976): to reduce the variation in student achievement and to reduce or eliminate any correlation between aptitude and achievement. Since all students must achieve at a high level on the subtraction objective but students who achieve the criterion early cannot go on to new material, there is a ceiling effect built into the procedure that will inherently cause variation among students to be small and will correspondingly reduce the correlation between mathematics aptitude and subtraction performance. In fact, if we were to set the mastery criterion at 100% and repeat the formative test-corrective instruction cycle until all students achieved this criterion, then the variance on the subtraction test would be zero, as would the correlation between aptitude and achievement.

However, this begs several critical questions. If some students take much longer than others to learn a particular objective, then one of two things must happen. Either corrective instruction must be given outside of regular class time, or students who achieve mastery early on will have to spend considerable amounts of time waiting for their classmates to catch up. The first option, extra time, is expensive and difficult to arrange, as it requires that teachers be available outside of class time to work with the nonmasters and that some students spend a great deal more time on any particular subject than they do ordinarily. The other option, giving enrichment or lateral extension activities to early masters while corrective instruction is given, may or may not be beneficial for these students. For all students mastery learning poses a dilemma, a choice between content coverage and content mastery (see Arlin, 1984a; Mueller, 1976; Resnick, 1977). It may often be the case that even for low achievers, spending the time to master each objective may be less productive than covering more objectives (see, for example, Cooley & Leinhardt, 1980).

Problems Inherent in Mastery Learning Research

The nature of mastery learning theory and practice creates thorny problems for research on the achievement effects of mastery learning strategies. These problems fall into two principal categories: unequal time and unequal objectives.

Unequal time. One of the fundamental propositions of mastery learning theory is that learning should be held constant and time should be allowed to vary, rather than the opposite situation held to exist in traditional instruction. However, if the total instructional time allocated to a particular subject is fixed, then a common level of learning for all students is likely to require taking time away from high achievers to increase it for low achievers, a leveling process that would in its extreme form be repugnant to most educators (see Arlin, 1982, 1984b; Arlin & Westbury, 1976; Fitzpatrick, 1985; Smith, 1981).

To avoid what Arlin (1984b) calls a "Robin Hood" approach to time allocation in mastery learning, many applications of mastery learning provide corrective instruction during times other than regular class time, such as during lunch, recess, or after school (see Arlin, 1982). In short-term laboratory studies, the extra time given to students who need corrective instruction is often substantial. For example, Arlin and Webster (1983) conducted an experiment in which students studied a unit on sailing under mastery or nonmastery conditions for 4 days. After taking formative tests, mastery learning students who did not achieve a score of 80% received individual tutoring during times other than regular class time. Nonmastery students took the formative tests as final quizzes and did not receive tutoring.

The mastery learning students achieved at twice the level of nonmastery students in terms of percent correct on daily chapter tests, an effect size (ES) of more than 3.0. However, mastery learning students spent more than twice as much time learning the same material. On a retention test taken 4 days after the last lesson, mastery students retained more than nonmastery students ($ES = .70$). However, nonmastery students retained far more *per hour of instruction* than did mastery learning students ($ES = -1.17$). Similarly, Gettinger (1985) found that students who were given enough time to achieve a 100% criterion on a set of reading tasks achieved only 15.5% more than did students who were allowed an average of half the time allocated to the 100% mastery group.

In recent articles published in *Educational Leadership* and the *Educational Researcher,* Benjamin Bloom (1984a, 1984b) noted that several dissertations done by his graduate students at the University of Chicago found effect sizes for mastery learning of 1 sigma or more (i.e., one standard deviation or more above the control group's mean). In all of these, corrective instruction was given outside of regular class time, increasing total instructional time beyond that allocated to the control groups. The additional time averaged 20–33% of the initial classroom instruction, or about 1 day per week. For example, in a 2-week study in Malaysia by Nordin (1979), an extra period for corrective instruction was provided to the mastery learning classes, while control classes did other school work unrelated to the units involved in the study. A 3-week study by Anania (1981) set aside one period each week for corrective instruction. In a study by Leyton (1983), students received 2–3 periods of corrective instruction for every 2–3 weeks of initial instruction.

In discussing the practicality of mastery learning, Bloom (1984a, p. 9) states that "the time or other costs of the mastery learning procedures have usually been very small." It may be true that school districts could in theory provide tutors to administer corrective instruction outside of regular class time; the costs of doing so would hardly be "very small," but cost or cost-effectiveness is not at issue here. But as a question of experimental design, the extra time often given to mastery learning classes is a serious problem. It is virtually unheard-of in educational research outside of the mastery learning tradition to systematically allocate an experimental group more instructional time than a control group, except in studies of the effects of time itself. Presumably, any sensible instructional program would produce significantly greater achievement than a control method that allocated 20–33% less instructional time. Studies that fail to hold time constant across treatments essentially confound treatment effects with effects of additional time.

It might be argued that mastery learning programs that provide corrective instruction outside of regular class time produce effects that are substantially greater *per unit time* than those associated with traditional instruction. However, computing "learning per unit time" is not a straightforward process. In the Arlin and Webster (1983) experiment discussed earlier, mastery learning students passed about twice as many items on immediate chapter tests as did control students, and the time allocated to the mastery learning students was twice that allocated to control. Thus, the "learning per unit time" was about equal in both groups. Yet on a *retention* test only 4 days later, the items passed per unit time were considerably higher for the control group. Which is the correct measure of learning per unit time, that associated with the chapter tests or that associated with the retention test?

Many mastery learning theorists (e.g., Block, 1972; Bloom, 1976; Guskey & Gates, 1985) have argued that the "extra time" issue is not as problematic as it seems, because the time needed for corrective instruction should diminish over time. The theory behind this is that by ensuring that all students have mastered the prerequisite skills for each new unit, the need for corrective instruction on each successive unit should be reduced. A few very brief experiments using specially constructed, hierarchically organized curriculum materials have demonstrated that over as many as three successive 1-hour units, time needed for corrective instruction does in fact diminish (Anderson, 1976; Arlin, 1973; Block, 1972). However, Arlin (1984a) examined time-to-mastery records for students involved in a mastery learning program over a 4-year period. In the first grade, the ratio of average time to mastery for the slowest 25% of students to that for the fastest 25% was 2.5 to 1. Rather than decreasing, as would have been predicted by mastery learning theorists, this ratio increased over the 4-year period. By the fourth grade, the ratio was 4.2 to 1. Thus, while it is theoretically possible that mastery learning procedures may ultimately reduce the need for corrective instruction, no evidence from long-term practical applications of mastery learning supports this possibility at present.

It should be noted that many studies of mastery learning do hold total instruction time more or less constant across experimental and control conditions. In discussing the "best evidence" on practical applications of mastery learning, issues of time for corrective instruction will be explored further.

Unequal objectives. An even thornier problem posed by research on mastery learning revolves around the question of achievement measures used as dependent variables. Most studies of mastery learning use experimenter-made summative achievement tests as the criterion of learning effects. The danger inherent in the use of such tests is that they will correspond more closely to the curriculum taught in the mastery learning classes than to that taught in control classes. Some articles on mastery learning experiments (e.g., Kersh, 1970; Lueckemeyer & Chiappetta, 1981) describe considerable efforts to ensure that experimental and control classes were pursuing the same objectives. Many studies administer the formative tests used in the mastery learning classes as quizzes in the control classes; in theory this should help focus the control classes on the same objectives. On the other hand, many other studies specified that students used the same texts and other materials but did not use formative tests in the control group or otherwise focus the control groups on the same objectives as those pursued in the mastery learning classes (e.g., Cabezon, 1984; Crotty, 1975).

The possibility that experimenter-made tests will be biased toward the objectives taught in experimental groups exists in all educational research that uses such tests, but it is particularly problematic in research on mastery learning, which by its nature focuses teachers and students on a narrow and explicitly defined set of objectives. When careful control of instruction methods, materials, and tests is not exercised, there is always a possibility that the control group is learning valuable information or skills not learned in the mastery learning group but not assessed on the experimenter-made measure.

Even when instructional objectives are carefully matched in experimental and control classes, use of experimenter-made tests keyed to what is taught in both classes can introduce a bias in favor of the mastery learning treatment. As noted earlier, when time for corrective instruction is provided during regular class time (rather than after class or after school), mastery learning trades *coverage* for *mastery* (see Anderson, 1985). The overall effects of this trade must be assessed using broadly based measures. What traditional whole-class instruction is best at, at least in theory, is covering material. Mastery learning proponents point out that material covered is not necessarily material learned. This is certainly true, but it is just as certainly true that material *not* covered is material *not* learned. Holding mastery learning and control groups to the same objectives in effect finesses the issue of instructional pace by measuring only the objectives covered by the mastery learning classes. If the control classes in fact cover more objectives, or could have done so had they not been held to the same pace as the mastery learning classes, this would not be registered on the experimenter-made test.

Two studies clearly illustrate the problems inherent in the use of experimenter-made tests to evaluate mastery learning. One is a year-long study of mastery learning in grades 1–6 by Anderson, Scott, and Hutlock (1976), which is described in detail later in this review. On experimenter-made math tests, the mastery learning classes significantly exceeded control at every grade level (mean effect size = +.64). On a retention test administered 3 months later, the experimental-control differences were still substantial (*ES* = +.49). However, the experimenters also used the mathematics scales from the standardized California Achievement Test as a dependent variable. On this test the experimental-control differences were effectively zero (*ES* = +.04).

A study by Taylor (1973) in ninth-grade algebra classes—although not strictly speaking a study of mastery learning—nevertheless illustrates the dilemma involved in the use of experimenter-made tests in evaluation of mastery learning programs. At the beginning of the semester, students in the experimental classes were each given a copy of a "minimal essential skills" test and were told that to pass the course they would need to obtain a score of at least 80% on a parallel form of the test. About 3 weeks before the end of the semester, another parallel form of the final test was administered to students, and the final 3 weeks was spent on remedial work and retesting for students who needed it (while other students worked on enrichment activities). At the end of the semester, the final test was given. A similar procedure was followed for the second semester.

Experimenter-made as well as standardized measures were used to assess the achievement effects of the program. On the minimum essential skills section of the experimenter-made test, scores averaged 87.3% correct, dramatically higher than they had been on the same test in the same schools the previous year (55.4%). On a section of the experimenter-made test covering skills "beyond, but closely related to, minimum essentials," differences favoring the experimental classes were still substantial, 44.6% correct versus 29.2%. Differences on the minimum essentials subtest of the standardized Cooperative Algebra Test also favored the experimental group (*ES* = +.47). However, on the section of the standardized test covering skills *beyond* minimum essentials, the control group exceeded the experimental group (*ES* = −.25).

The Taylor (1973) intervention does not qualify as mastery learning because it involved only one feedback-corrective instruction cycle per semester. However, the study demonstrates a problem characteristic of mastery learning studies that use experimenter-made tests as dependent measures. Had Taylor used only the experimenter-made test, his study would have appeared to provide overwhelming support for the experimental procedures. However, the results for the standardized tests indicated that students in the control group (the previous year) were learning materials that did not appear on the experimenter-made tests. The attention and efforts of teachers as well as students were focused on a narrow set of instructional objectives that constituted only about 30% of the items on the broader-based standardized measure.

These observations concerning problems in the use of experimenter-made measures do not imply that all studies that use them should be ignored. Rather, they are meant to suggest extreme caution and careful reading of details of each such study before conclusions are drawn.

Methods

This review uses a method called "best-evidence synthesis," procedures described by Slavin (1986a) for synthesizing large literatures in social science. Best-evidence synthesis essentially combines the quantification of effect sizes and the systematic literature search and inclusion procedures of meta-analysis (Glass et al., 1981) with the description of individual studies and methodological and substantive issues characteristic of traditional literature reviews. In order to allow for adequate description of a set of studies high in internal and external validity, best-evidence synthesis applies well-justified a priori criteria to select studies to constitute the main body of the review (see Slavin, 1986b, for an earlier example of this procedure).

This section, "Methods," outlines the specific procedures used in preparing the review, including such issues as how studies were located, which were selected for inclusion, how effect sizes were computed, how studies were categorized, and how the question of pooling of effect sizes was handled.

Literature Search Procedures

The first step in conducting the best-evidence synthesis was to locate as complete as possible a set of studies of mastery learning. Several sources of references were used. The ERIC system and Dissertation Abstracts produced hundreds of citations in response to the key words "mastery learning." Additional sources of citations included a bibliography of mastery learning studies compiled by Hymel (1982), earlier reviews and meta-analyses on mastery learning, and references in the primary studies. Papers presented at the American Educational Research Association meetings since 1976 were solicited from their authors. Dissertations were ordered from University Microfilms and from the University of Chicago, which does not cooperate with University Microfilms.

Criteria for Study Inclusion

The studies on which this review is primarily based had to meet a set of a priori criteria with respect to germaneness and methodological adequacy.

Germaneness. To be considered germane to the review, all studies had to evaluate group-based mastery learning programs in regular (i.e., nonspecial) elementary and secondary classrooms."Group-based mastery learning" was defined as any instructional method that had the following characteristics:

1. Students were tested on their mastery of instructional objectives at least once every 4 weeks. A mastery criterion was set (e.g., 80% correct), and students who did not achieve this criterion on an initial formative test received corrective instruction and a second formative or summative test. This cycle could be repeated one or more times. Studies were included regardless of the form of corrective instruction used and regardless of whether corrective instruction was given during or outside of regular class time.

2. Before each formative test, students were taught as a total group. This requirement excluded studies of individualized or continuous progress forms of mastery learning and studies of the Personalized System of Instruction. However, studies in which mastery learning students worked on individualized materials as corrective (not initial) instruction were included.

3. Mastery learning was the only or principal intervention. This excluded comparisons such as those in two studies by Mevarech (1985a, 1985b) evaluating a combination of mastery learning and cooperative learning, and comparisons involving enhancement of cognitive entry behaviors (e.g., Leyton, 1983).

Studies evaluating programs similar to mastery learning but conducted before Bloom (1968) described it were excluded (e.g., Rankin, Anderson, & Bergman, 1936). Other than this, no restrictions were placed on sources or types of publications. Every attempt was made to locate dissertations, ERIC documents, and conference papers as well as published materials.

Methodological Adequacy. Criteria for methodological adequacy were as follows.

1. Studies had to compare group-based mastery learning programs to traditional group-paced instruction not using the feedback-corrective cycle. A small number of studies (e.g., Katims & Jones, 1985; Strasler & Rochester, 1982) that compared achievement under mastery learning to that during previous years (before mastery learning was introduced) were excluded, on the basis that changes in grade-to-grade promotion policies, curriculum alignment, and other trends in recent years make year-to-year changes difficult to ascribe to any one factor.

2. Evidence had to be given that experimental and control groups were initially equivalent, or the degree of nonequivalence had to be quantified and capable of being adjusted for in computing effect sizes. This excluded a small number of studies which failed either to give pretests or to randomly assign students to treatments (e.g., Dillashaw & Okey, 1983).

3. Study duration had to be at least 4 weeks (20 hours). This restriction excluded a large number of brief experiments that often used procedures that would be difficult to replicate in practice (such as providing 1 hour of corrective instruction for every hour of initial instruction). The reason for this restriction was to concentrate the review on mastery learning procedures that could in principle be used over extended time periods. One 4-week study by Strasler (1979) was excluded on the basis that it was really two 2-week studies on two completely unrelated topics, ecology and geometry. The 4-week requirement caused by far the largest amount of exclusion of studies included in previous reviews and meta-analyses. For example, of 25 elementary and secondary achievement studies cited by Guskey and Gates (1985), 11 (with a median duration of 1 week) were excluded by this requirement. However, it should be noted that most of these brief studies would also have been excluded by other criteria, principally use of individualized rather than group-based forms of mastery learning and inclusion of only one class per treatment (see below).

4. At least two experimental and two control classes and/or teachers had to be involved in the study. This excluded a few studies (e.g., Collins, 1971; Leyton, 1983; Long, Okey, & Yeany, 1981; Mevarech, 1985a; Tenenbaum, 1982) in which, because only one teacher taught in each treatment condition, treatment effects were completely confounded with teacher/class effects. Also excluded were a few studies in which several teachers were involved but each taught a different subject

(Guskey, 1982, 1984; Okey, 1974, 1977; Rubovits, 1975). Because it would be inappropriate to compute effect sizes across the different subjects, these studies were seen as a set of two-class comparisons, each of which confounded teacher and class effects with treatment effects.

5. The achievement measure used had to be an assessment of objectives taught in control as well as experimental classes. This requirement was liberally interpreted and excluded only one study, a dissertation by Froemel (1980) in which the mastery learning classes' summative tests were used as the criterion of treatment effects and no apparent attempt was made to see that the control classes were pursuing the same objectives. In cases in which it was unclear to what degree control classes were held to the same objectives as experimental classes and experimenter-made measures were used, the studies were included. These studies are identified and discussed later in this review, and their results should be interpreted with a great deal of caution.

Also excluded were studies that used grades as the only dependent measures (e.g., Mathews, 1982; Wortham, 1980). In group-based mastery learning, grades are increased as part of the treatment, as students have opportunities to take tests over to try to improve their scores. They are thus not appropriate as measures of the achievement effects of the program. Similarly, studies that used time on-task as the only dependent measure were excluded (e.g., Fitzpatrick, 1985).

Computation of Effect Sizes

The size and direction of effects of mastery learning on student achievement are presented throughout this review in terms of effect size. Effect size, as described by Glass et al. (1981), is the difference between experimental and control posttest means divided by the control group's posttest standard deviation. However, this formula was adapted in the present review to take into account pretest or ability differences between the experimental and control groups. If pretests were available, then the formula used was the difference in experimental and control *gains* divided by the control group's posttest standard deviation. If ability measures rather than pretests were presented, then the experimental-control difference on these measures, divided by the control group's standard deviation, was subtracted from the posttest effect size (this was necessary in only one case, a study by Cabezon, 1984). The reason for these adjustments is that in studies of achievement, posttest scores are so highly correlated with pretest levels that any pretest differences are likely to be reflected in posttests, correspondingly inflating or deflating effect sizes computed on posttests alone. These adjustments are not precisely those recommended by Glass et al. (1981), who present formulas for dealing with gain scores that rely on knowledge of pre-post correlations (which are rarely reported). However, the adjustment procedures used in the present paper follow Glass et al. (1981) in accounting for pretest differences while preserving the control group's standard deviation as the common metric of effect size. Such procedures as ignoring pretest information or using standard deviations of gain scores as the denominator in computing effect sizes are often seen in meta-analyses but are explicitly rejected by Glass and his colleagues (see Glass et al., 1981, pp. 115–119).

Because individual-level standard deviations are usually of concern in mastery learning research, most studies that met other criteria for inclusion presented data sufficient for direct computation of effect size. In many studies, data analyses used *class* means and standard deviations, but individual-level standard deviations were also presented. In every case (following Glass et al., 1981) the individual-level standard deviations were used to compute effect sizes; class-level standard deviations are usually much smaller than individual-level *SD*s, inflating effect size estimates. Also, note that the control group standard deviation, not a pooled standard deviation, was always used, as mastery learning often has the effect of reducing achievement standard deviations.

In the few cases in which data necessary for computing effect sizes were lacking in studies which otherwise met criteria for inclusion, the studies' results were indicated in terms of their direction and statistical significance.

Research on Achievement Effects of Group-Based Mastery Learning

What are the effects of group-based mastery learning on the achievement of elementary and secondary students? In essence, there are three claims that proponents of mastery learning might make for the effectiveness of mastery learning. These are as follows:

1. Mastery learning is more effective than traditional instruction *even when instructional time is held constant and achievement measures register coverage as well as mastery.* This might be called the "strong claim" for the achievement effects of mastery learning. It is clear, at least in theory, that if mastery learning procedures greatly increase allocated time for instruction by providing enough additional time for corrective instruction to bring all students to a high level of mastery, then mastery learning students will achieve more than traditionally taught control students. But it is less obviously true that the additional time for corrective instruction is more productive in terms of student achievement than it would be simply to increase allocated time for the control students. The "strong claim" asserts that time used for corrective instruction (along with the other elements of mastery learning) is indeed more productive than time used for additional instruction to the class as a whole. It is important to note that this "strong claim" might not be endorsed by all mastery learning proponents. For example, Bloom (1976, p. 5) notes that the "time costs [necessary to enable four fifths of students to reach a level of achievement that less than one fifth attain in nonmastery conditions] are typically of the order of ten to twenty percent additional time over the classroom scheduled time." However, Block and Anderson (1975) describe a form of mastery learning that can be implemented within usual time constraints, and in practice corrective instruction is rarely given during additional time.

Similarly, it is clear (in theory) that if students who experienced mastery learning are tested on the specific objectives they studied, they will score higher on those objectives than will students who were studying similar but not identical objectives. Further, it is likely that even if mastery learning and control classes are held to precisely the same objectives but the control classes are not allowed to move ahead if they finish those objectives before their mastery learning counterparts do, then the traditional model is deprived of its natural advantage, the capacity to cover material rapidly. A "fair" measure of student achievement in a mastery learning experiment would have to register both coverage and mastery, so that if the control group covered more objectives than the mastery learning group its learning of these additional objectives would be registered. The "strong claim" would hold that, even allowing control classes to proceed at their own rate and even using such an achievement measure, mastery learning would produce more achievement than control methods.

The best evidence for the "strong claim" would probably come from studies in which mastery learning and control classes studied precisely the same objectives using the same materials and lessons and the same amount of allocated time, but in which teachers could determine their own pace of instruction and achievement measures covered the objectives reached by the fastest-moving class. Unfortunately, such studies are not known to exist. However, a good approximation of these experimental design features is achieved by studies that hold allocated time constant and use standardized tests as the criterion of achievement. Assuming that curriculum materials are not specifically keyed to the standardized tests in either treatment, these tests offer a means of registering both mastery and coverage. In such basic skills areas as mathematics and reading, the standardized tests are likely to have a high overlap with the objectives pursued by mastery learning teachers as well as by control teachers.

2. Mastery learning is an effective means of ensuring that teachers adhere to a particular curriculum and students learn a specific set of objectives (the "curricular focus" claim). A "weak claim" for the effectiveness of mastery learning would be that these methods focus teachers on a particular set of objectives held to be superior to those that might have been pursued by teachers on their own. This might be called the "curricular focus" claim. For example, consider a survey course on U.S. history. Left to their own devices, some teachers might teach details about individual battles of the Civil War; others might entirely ignore the battles and focus on the economic and political issues; and still others might approach the topic in some third way, combine both approaches, or even teach with no particular plan of action. A panel of curriculum experts might determine that there is a small set of critical understandings about the Civil War that all students should have, and they might devise a criterion-referenced test to assess these understandings. If it can be assumed that the experts' judgments are indeed superior to those of individual teachers, then teaching to this test may not be inappropriate, and mastery learning may be a means of holding students and teachers to the essentials, relegating other concepts they might have learned (that are not on the criterion-referenced test) to a marginal status. It is no accident that mastery learning grew out of the behavioral objectives/criterion-referenced testing movement (see Bloom, Hastings, & Madaus, 1971); one of the central precepts of mastery learning is that once critical objectives are identified for a given course, students should be required to master those and only those objectives. Further, it is interesting to note that in recent years the mastery learning movement has often allied itself with the "curriculum alignment" movement, which seeks to focus teachers on objectives that happen to be contained in district- and/or state-level criterion-referenced minimum competency tests as well as norm-referenced standardized tests (see Levine, 1985).

The "curricular focus" claim, that mastery learning may help focus teachers and students on certain objectives, is characterized here as a "weak claim" because it requires a belief that any objectives other than those pursued by the mastery learning program are of little value. Critics (e.g., Resnick, 1977) point out with some justification that a focus on a well-defined set of minimum objectives may place a restriction on the maximum that students might have achieved. However, in certain circumstances it may well be justifiable to hold certain objectives to be essential to a course of study, and mastery learning may represent an effective means of ensuring that nearly all students have attained these objectives.

The best evidence for the "curricular focus" claim would come from studies in which curriculum experts formulated a common set of objectives to be pursued equally by mastery learning and control teachers within an equal amount of allocated time. If achievement on the criterion-referenced assessments were higher in mastery learning than in control classes, we could at least make the argument that the mastery learning students have learned more of the *essential* objectives, even though the control group may have learned additional, presumably less essential concepts.

3. Mastery learning is an effective use of additional time and instructional resources to bring almost all students to an acceptable level of achievement (the "extra time" claim). A second "weak claim" would be that given the availability of additional teacher and student time for corrective instruction, mastery learning is an effective means of ensuring a minimal level of achievement for all students. As noted earlier, in an extreme form this "extra time" claim is almost axiomatically true. Leaving aside cases of serious learning disabilities, it should certainly be possible to ensure that virtually all students can achieve a minimal set of objectives in a new course if an indefinite amount of one-to-one tutoring is available to students who initially fail to pass formative tests. However, it may be that, even within the context of the practicable, providing students with additional instruction if they need it will bring almost all to a reasonable level of achievement.

The reason that this is characterized here as a "weak claim" is that it begs the question of whether the additional time used for corrective instruction is the *best* use of additional time. What could the control classes do if they also had more instructional time? However, the "extra time" issue is not a trivial one, as it is not impossible to routinely provide corrective instruction to students who need it outside of regular class time. For example, this might be an effective use of compensatory (Chapter I) or special education resource pullouts, a possibility that is discussed later.

The best evidence for this claim would come from studies that provided mastery learning classes with additional time for corrective instruction and used achievement tests that covered all topics that could have been studied by the fastest-paced classes (e.g., standardized tests). However, such studies are not known to exist; the best existing evidence for the "extra time" claim is from studies that used experimenter-made achievement measures and provided corrective instruction outside of class time.

Evidence for the "Strong Claim"

Table 1 summarizes the major characteristics and findings of seven mastery learning studies that met the inclusion criteria discussed earlier, provided equal time for experimental and control classes, and used standardized measures of achievement.

Table 1 clearly indicates that the effects of mastery learning on standardized achievement measures are extremely small, at best. The median effect size across all seven studies is essentially zero ($ES = +.04$). The only study with a nontrivial effect size ($ES = +.25$), a semester-long experiment in inner-city Chicago elementary schools by Katims, Smith, Steele, & Wick (1977), also had a serious design

TABLE 1

Equal-time studies using standardized measures

Article	Grades	Location	Sample size	Duration	Design	Treatments	Subjects	Effect sizes: By group/measure	Effect sizes: Total
Elementary									
Anderson et al., 1976	1–6	Lorain, OH	2 sch.	1 yr.	Students in matched ML[a], control schools matched on ability	ML—Followed Block (1971). Control—Untreated	Math		+.04
Kersh, 1970	5	Suburban Chicago	11 cl.	1 yr.	Teachers/classes randomly assigned to ML, control within each school	ML—Corr. inst. included reteaching, alternative mtls., peer tutoring; formative tests given every 3–4 wks. Control—Untreated	Math	middle cl. (−) lower cl. (+)	0
Gutkin, 1985	1	Inner-city New York	41 cl.	1 yr.	Schools randomly assigned to ML, control	ML—Formative tests given every month Control—Untreated	Reading		+.12
Katims et al., 1977	Upper elem.	Inner-city Chicago	19 cl.	15 wks.	1 ML, 1 cont. class from each of 10 schools. Trts. self-selected or principal-imposed	ML—specific mtls. provided Control—Untreated	Reading		+.25
Jones et al., 1979	Upper elem.	Inner-city Chicago	4 sch.	1 yr.	2 ML schools matched with 2 control schools	ML—specific mtls. provided Control—Untreated	Reading		+.09
Secondary									
Slavin & Karweit, 1984	9	Inner-city Philadelphia	25 cl.	26 wks.	Teachers/classes randomly assigned to ML, control	ML—Formative tests given every 2–3 wks.; Corr. inst. given by teachers. Control—Used same mtls., tests, procedures as ML except for corr. inst. & summative tests	General math	Hi 0 Lo 0	+.02
Chance, 1980	8	Inner-city New Orleans	6 cl.	5 wks.	Students within each of 3 classes randomly assigned to ML or control	ML—Formative tests given every wk. Mast. crit. = 80–90% Control—Used same mtls., tests, procedures as ML	Reading	Hi 0 Av 0 Lo 0	0

[a] Mastery learning
(+) Nonsignificant difference favoring ML
0 No difference
(−) Nonsignificant difference favoring control

flaw. Teachers were allowed to select themselves into mastery learning or control treatments or were assigned to conditions by their principals. It is entirely possible that the teachers who were most interested in using the new methods and materials, or those who were named by their principals to use the new program, were better teachers than were the control teachers. In any case, the differences were not statistically significant when analyzed at the class level, were only marginally significant ($p = .071$) for individual-level gains, and amounted to an experimental-control difference of only 11% of a grade equivalent.

The Katims et al. (1977) study used a specially developed set of materials and procedures that became known as the Chicago Mastery Learning Reading program, or CMLR. This program provides teachers with specific instructional guides, worksheets, formative tests, corrective activities, and extension materials. A second study of CMLR by Jones, Monsaas, and Katims (1979) compared matched CMLR and control schools over a full year. This study found a difference between CMLR and control students on the Iowa Test of Basic Skills Reading Comprehension scale that was marginally significant at the individual level but quite small ($ES = +.09$). In contrast, on experimenter-made "end-of-cycle" tests the mastery learning classes did significantly exceed control ($ES = +.18$). A third study of CMLR by Katims and Jones (1985) did not qualify for inclusion in Table 1 because it compared year-to-year gains in grade equivalents rather than comparing experimental to control groups. However, it is interesting to note that the difference in achievement gains between the cohort of students who used the CMLR program and those in the previous year who did not was only 0.16 grade equivalents, which is similar to the results found in the Katims et al. (1977) and Jones et al. experimental-control comparisons.

One of the most important studies of mastery learning is the year-long Anderson, Scott, and Hutlock (1976) experiment briefly described earlier. This study compared students in grades 1–6 in one mastery learning and one control school in Lorain, Ohio. The school populations were similar, but there were significant pretest differences at the first- and fourth-grade levels favoring the control group. To ensure initial equality in this nonrandomized design, students were individually matched on the Metropolitan Readiness Test (grades 1–3) or the Otis-Lennon Intelligence Test (grades 4–6). In the mastery learning school, students experienced the form of mastery learning described by Block and Anderson (1975). The teachers presented a lesson to the class and then assessed student progress on specific objectives. "Errors . . . were remediated through the use of both large-group and small-group re-learning and review sessions. After every student had demonstrated mastery on the formative test for each unit, the class moved on to the next unit" (Anderson et al., 1976, p. 4).

One particularly important aspect of the Anderson et al. (1976) study is that it used both standardized tests and experimenter-made, criterion-referenced tests. The standardized tests were the Computations, Concepts, and Problem Solving scales of the California Achievement Test. The experimenter-made test was constructed by the project director (Nicholas Hutlock) to match the objectives taught in the mastery learning classes. Control teachers were asked to examine the list of objectives and identify any they did not teach, and these were eliminated from the test.

The results of the study were completely different for the two types of achievement tests. On the experimenter-made tests, students in the mastery learning classes achieved significantly more than did their matched counterparts at every grade level (mean *ES* = +.64). A retention test based on the same objectives was given 3 months after the end of the intervention period, and mastery learning classes still significantly exceeded control (*ES* = +.49). However, on the standardized tests, these differences were not registered. Mastery learning students scored somewhat higher than control on Computations (*ES* = +.17) and Problem Solving (*ES* = +.07), but the control group scored higher on Concepts (*ES* = −.12).

The Anderson et al. (1976) finding of marked differences in effects on standardized and experimenter-made measures counsels great caution in interpreting results of other studies that used experimenter-made measures only. In a year-long study of mathematics, it is highly unlikely that a standardized mathematics test would fail to register any meaningful treatment effect. Therefore, it must be assumed that the strong positive effects found by Anderson et al. on the experimenter-made tests are mostly or entirely due to the fact that these tests were keyed to the mastery learning classes' objectives. It may be that the control classes covered more objectives than the mastery learning classes, and that learning of these additional objectives was registered on the standardized but not the experimenter-made measures.

Another important study of mastery learning at the elementary level is a dissertation by Kersh (1970), in which 11 fifth-grade classes were randomly assigned to mastery learning or control conditions for an entire school year. Two schools were involved, one middle-class and one lower class. Students' math achievement was assessed about once each month in the mastery learning classes, and peer tutoring, games, and other alternative activities were provided to students who did not show evidence of mastery. Control classes were untreated. The study results did not favor either treatment overall on the Stanford Achievement Test's Concepts and Applications scales. Individual-level effect sizes could not be computed, as only class-level means and standard deviations were reported. However, class-level effect sizes were essentially zero in any case (*ES* = −.06). On an experimenter-made criterion-referenced test not specifically keyed to the mastery objectives, the results were no more conclusive; class-level effects slightly favored the control group (*ES* = −.20). Effects somewhat favored mastery learning in the lower class school and favored the control group in the middle-class school, but since none of the differences approached statistical significance these trends may just reflect teacher effects or random variation.

In a recent study by Gutkin (1985), 41 first-grade classes in New York City were randomly assigned to mastery learning or control treatments. The article does not describe the mastery learning treatment in detail, except to note that monthly formative tests were given to assess student progress through prescribed instructional units. The mastery learning training also included information on classroom management skills, process-product research, and performance-based teacher education, and teachers received extensive coaching, routine feedback from teacher trainers, and scoring services for formative and summative tests. After one year, mastery learning-control differences did not approach statistical significance in Total Reading on the California Achievement Test (*ES* = +.12). However, effects were more positive on a Phonics subscale (*ES* = +.36) than on Reading Vocabulary (*ES* = +.04) or Reading Comprehension (*ES* = +.15). Phonics, with its easily measurable objectives, may lend itself better to the mastery learning approach than do reading comprehension or vocabulary.

Studies using standardized measures at the secondary level are no more supportive of the "strong claim" than are the elementary studies. A 26-week experiment in inner-city, mostly black, Philadelphia junior and senior high schools assessed mastery learning in ninth grade "consumer mathematics," a course provided for students who do not qualify for Algebra I (Slavin & Karweit, 1984). Twenty-five teachers were randomly assigned to mastery learning or control treatments, both of which used the same books, worksheets, and quizzes in the same cycle of activities. However, instructional pace was not held constant. After each 1-week unit (approximately), mastery learning classes took a formative test, and then any students who did not achieve a score of at least 80% received corrective instruction from the teacher while those who did achieve at that level did enrichment activities. The formative tests were used as quizzes in the control group, and after taking the quizzes the class went on to the next unit.

Results on a shortened version of the Comprehensive Test of Basic Skills Computations and Concepts and Applications scales indicated no differences between mastery learning and control treatments ($ES = +.02$), and no interaction with pretest level; neither low nor high achievers benefited from the mastery learning model. It is interesting to note that there were two other treatment conditions evaluated in this study, a cooperative learning method called Student Teams–Achievement Divisions (STAD) (Slavin, 1983), and a combination of STAD and mastery learning. STAD classes did achieve significantly more than control ($ES = +.19$), but adding the mastery learning component to STAD had little additional achievement effect ($ES = +.03$).

A 5-week study by Chance (1980) compared randomly assigned mastery learning and control methods in teaching reading to students in an all-black, inner-city New Orleans school. Approximately once each week, students in the mastery learning groups took formative tests on unit objectives. If they did not achieve at 80% on three quizzes or 90% on one, they received tutoring, games, and/or manipulatives to correct their errors and had three opportunities to pass. No effects for students at any level of prior performance were found on the Gates-McGinitie Comprehension Test. However, it may be unrealistic to expect effects on a standardized measure after only 5 weeks.

Overall, research on the effects of mastery learning on standardized achievement test scores provides little support for the "strong claim" that, holding time and objectives constant, mastery learning will accelerate student achievement. The studies assessing these effects are not perfect; particularly when mastery learning is applied on a fairly wide scale in depressed inner-city schools, there is reason to question the degree to which the model was faithfully implemented. However, most of the studies used random assignment of classes or students to treatments, study durations approaching a full school year, and measures that registered coverage as well as mastery. Not one of the seven studies found effects of mastery learning that reached even conventional levels of statistical significance (even in individual-level analyses), much less educational significance. If group-based mastery learning had strong effects on achievement in such basic skills as reading and math, these studies would surely have detected them.

Evidence for the "Curricular Focus" Claim

Table 2 summarizes the principal evidence for the "curricular focus" claim, that mastery learning is an effective means of increasing student achievement of *specific* skills or concepts held to be the critical objectives of a course of study. The studies listed in the table are those that (in addition to meeting general inclusion criteria) used experimenter-made, criterion-referenced measures and apparently provided experimental and control classes with equal amounts of instructional time. It is important to note that the distinction between the equal-time studies listed in Table 2 and the unequal-time studies in Table 3 is often subtle and difficult to discriminate, as many authors did not clarify when or how corrective instruction was delivered or what the control groups were doing during the time when mastery learning classes received corrective instruction.

A total of nine studies met the requirements for inclusion in Table 2. Three of these (Anderson et al., 1976; Jones et al., 1979; Kersh, 1970) were studies that used both standardized and experimenter-made measures and were therefore also included in Table 1 and discussed earlier.

All but one (Kersh, 1970) of the studies listed in Table 2 found positive effects of mastery learning on achievement of specified objectives, with five studies falling in an effect size range from +.18 to +.27. The overall median effect size for the eight studies that used immediate posttests is +.255. However, the studies vary widely in duration, experimental and control treatments, and other features, so this median value should be cautiously interpreted.

Fuchs, Tindal, and Fuchs (1985) conducted a small and somewhat unusual study of mastery learning in rural first-grade reading classes. Students in four classes were randomly assigned to one of two treatments. In the mastery learning classes, students were tested on oral reading passages in their reading groups each week. The whole reading group reviewed each passage until at least 80% of the students could read the passage correctly at 50 words per minute. The control treatment was held to be the form of "mastery learning" recommended by basal publishers. These students were given unit tests every 4–6 weeks, but all students went on to the next unit regardless of score. Surprisingly, the measure on which mastery learning classes exceeded control was "end-of-book" tests provided with the basals ($ES = +.35$), not passage reading scores that should have been more closely related to the mastery learning procedures ($ES = +.05$). On both measures it was found that while low achievers benefited from the mastery learning approach, high achievers generally achieved more in the control classes. Since the control teachers were presumably directing their efforts toward the objectives assessed in the end-of-book tests to the same degree as the mastery learning teachers, the results on this measure are probably fair measures of achievement. However, the Fuchs et al. study may be more a study of the effects of repeated reading than of mastery learning per se. Research on repeated reading (e.g., Dahl, 1979) has found this practice to increase comprehension of text.

Another small and unusual study at the elementary level was reported by Wyckoff (1974), who randomly assigned four sixth-grade classes to experimental or control conditions for a 9-week anthropology unit. Following teaching of each major objective, students were quizzed. If the class median was at least 70% correct, the class moved on to the next objective; otherwise, those who scored less than 70%

received peer tutoring or were given additional reading or exercises. The control groups used precisely the same materials, tests, and schedule. The achievement results were not statistically significant, but they favored the mastery learning classes (*ES* = +.24). However, this trend was entirely due to effects on low performance readers (*ES* = +.58), not high-ability readers (*ES* = +.03).

One remarkable study spanning grades 3, 6, and 8 was reported in a dissertation by Cabezon (1984). The author, the director of the National Center for Curriculum Development in Chile, was charged with implementation of mastery learning throughout that country. Forty-one elementary schools throughout Chile were selected to serve as pilots, and an additional 2,143 *schools* began using mastery learning 2 years later. Three years after the pilots had begun, Cabezon randomly selected a sample of schools that had been using mastery learning for 3 years, for 1 year, or not at all. Within each school two classes at the third-, sixth-, and eighth-grade levels were selected.

The form of mastery learning used was not clearly specified, but teachers were expected to assess student progress every 2–3 weeks and to provide corrective instruction to those who needed it. Two subjects were involved, Spanish and mathematics.

Unfortunately, the classes that had used mastery learning for 3 years were found to be much higher in socioeconomic status (SES) and mean IQ level than were control classes. Because of this problem these comparisons did not meet the inclusion criteria. However, the classes that had used mastery learning for 1 year were comparable to the control classes in SES and only slightly higher in IQ.

The study results, summarized in Table 2, indicated stronger effects of mastery learning in Spanish than in math, and stronger effects in the early grades than in later ones, with an overall mean of +0.27. However, while all teachers used the same books, it is unclear to what degree control teachers were held to or even aware of the objectives being pursued by the mastery learning schools.

Two studies at the secondary level assessed both immediate and long-term impacts of mastery learning. One was a study by Lueckemeyer and Chiappetta (1981), who randomly assigned 10th graders to six mastery learning or six control classes for a 6-week human physiology unit. In the mastery learning classes, students were given a formative test every 2 weeks. They were then given 2 days to complete corrective activities for any objectives on which they did not achieve an 80% score, following which they took a second form of the test, which was used for grading purposes. Students who achieved the 80% criterion on the first test were given material to read or games to play while their classmates received corrective instruction. The control group studied the same material and took the same tests, but did not receive the 2-day corrective sessions. The control teachers were asked to complete the three 2-week units in 6 weeks, but were not held to the same schedule as the mastery classes. In order to have time to fit in the 2 days for corrective instruction every 2 weeks, the mastery learning classes "had to condense instruction. . .and to guard carefully against any wasted time (C. L. Lueckemeyer, personal communication, November 4, 1986).

On an immediate posttest the mastery learning classes achieved significantly more than the control group (*ES* = +.39), but on a retention test given 4 weeks later the difference had disappeared. The study's authors reported the statistically significant effects on posttest achievement but noted that "it is questionable whether

TABLE 2
Equal-time studies using experimenter-made measures

Article	Grades	Location	Sample size	Duration	Design	Treatments	Subjects	Effect sizes: By group/measure	Effect sizes: Total	Retention
Elementary										
Anderson et al., 1976	1–6	Lorain, OH	2 sch.	1 yr.	(See Table 1)	(See Table 1)	Math	Posttest Retention (3 mo.)	+.64	+.49
Kersh, 1970	5	Suburban Chicago	11 cl.	1 yr.	(See Table 1)	(See Table 1)	Math		(−)	
Jones et al., 1979	Upper elem.	Inner-city Chicago	4 sch.	1 yr.	(See Table 1)	(See Table 1)	Reading		+.18	
Wyckoff, 1974	6	Suburban Atlanta	4 cl.	9 wks.	Teachers/classes randomly assigned to ML[a], control	ML—Mastery criterion 70%. Corr. inst. was either reteaching to whole class or peer tutoring Control—Used same mtls., tests, procedures as ML except for corr. inst. and summative tests	Anthropology	Hi +.03 Lo +.58	+.24	
Fuchs et al., 1985	1	Rural Minnesota	4 cl.	1 yr.	Students randomly assigned to ML, control	ML—Students tested on oral rdg. passages each wk. Whole rdg. grp. reviewed until 80% of students got at least 50 wpm correct Control—students tested every 4–6 wks., all were promoted w/o corr. inst.	Reading	Hi (−) Lo +	+.20	
Elementary and secondary										
Cabezon, 1984	3, 6, 8	Chile	46 cl.	1 yr.	Compared classes using ML to classes similar in SES, IQ	ML—Not clearly specified Control—Untreated	Spanish +.40 Math +.14	Gr3 + .47 Gr6 + .22 Gr8 + .12	+.27	

Secondary										
Lueckemeyer & Chiappetta, 1981	10	Suburban Houston	12 cl.	6 wks.	Students randomly assigned to ML, control Pretest differences favored control	ML—Formative tests given every 2 wks., followed 2 days of corr. inst. (criterion = 80%)	Human Physiology	Posttest	+.39	
						Control—Used same mtls., tests, procedures as ML except for corr. inst. and summative tests		Retention (4 wks.)		0
Dunkelberger & Heikkinen, 1984	9	Suburban Delaware	10 cl.	15 wks.	Students randomly assigned to ML, control classes. Teachers taught ML & control classes. Posttest given > 4 mos. after end of implementation period	ML—Students had to meet 80% criterion on repeatable tests to go on. Corrective activities available during free time Control—Used same mtls., procedures, tests. Received detailed feedback and had same corrective mtls. available during free time	Chem., Physics	Retention (4 mo.)		+.26
Mevarech, 1986	7	Israel	4 cl.	3 mo.	Students randomly assigned to ML, control classes	ML—Students who did not reach 70% crit. on formative tests rec'd corr. inst. from tch. or peers Control—Used same mtls., tests as ML	Algebra	Lo SES +1.78 Mid SES +.91 Hi SES +.66	+.90	

[a] Mastery learning
+ Significant difference favoring ML
0 No difference
(−) Nonsignificant difference favoring control

TABLE 3
Unequal-time studies using experimenter-made measures (secondary)

Article	Grades	Location	Sample size	Duration	Design	Treatments	Extra time	Subjects	Effect sizes: By group/measure	Effect sizes: Total	Retention
Long et al., 1978	8	Georgia	6 cl.	5 wks.	Students randomly assigned to 3 trts. Teachers rotated across trts.	Teacher-directed ML[a]—Formative tests given every 2 days. Remedial work given as corr. inst. "If problem persists," indiv. tutoring given by teacher Student-directed ML—Same formative tests used, returned to students for self-correction Control—Same inst. but no tests, correctives	Not stated	Earth Sci.	Teacher-directed ML vs. Control: Posttest Retention (12 wks.) Teacher-directed ML vs. Student-directed ML: Posttest Retention (12 wks.)	 +.43 +.19	 +.08 −.03
Fagan, 1976	7	Dallas, TX, 1 middle class sch., 1 lower class sch.	17 cl.	5 wks.	4 teachers randomly assigned to ML, control	ML—Formative tests given every wk. Teachers drilled students who failed to achieve 80% criterion, then gave 2nd formative test. Control—Used same mtls. & procedures as ML. Formative tests taken as quizzes	22%	Transp. and Environ.	Posttest Retention (4 wks.)	−.11	 −.15
Hecht, 1980	10	Urban, suburban Midwest	5 cl.	6 wks.	Students randomly assigned to ML, control classes. Two teachers taught ML & control classes.	ML—Formative tests given every 2 wks., followed by "intensive remedial help" Control—Used same mtls. & procedures as ML including both 1st & 2nd formative tests, but no remedial help	Not stated	Geometry		+.31	

Mevarech, 1980	9	Chicago, middle class sch.	8 cl.	6 wks.	Students randomly assigned in 2 × 2 design to "algorithmic strategy" vs. "heuristic strategy" and to ML vs. control	ML—Formative tests given every 2 wks. Students had 3 chances to obtain 80% criterion. Corr. inst. included grp. inst., peer tutoring, adult tutoring outside of class. Control—Used same mtls. & procedures, took formative tests as quizzes. While ML classes received corr. inst., control worked add'l problems.	Not stated	Alg. I	Algorithmic Strategy +.70 Heuristic Strategy +.83	+.77

[a] Mastery learning.

such a limited effect on achievement is worth the considerable time required for the development and management of such an instructional program" (Lueckemeyer & Chiappetta, 1981, p. 273). Further, it is unclear whether the control groups required the full 6 weeks to cover the material. Any additional information students in the control group learned (or could have learned) would of course not have been registered on the experimenter-made test.

In a 15-week experiment in ninth-grade chemistry and physics classes by Dunkelberger and Heikkinen (1984), students were randomly assigned to mastery learning or control classes. In the mastery learning classes, students had several chances to meet an 80% criterion on parallel formative tests. Control students took the tests once and received feedback on their areas of strength and weakness. All students, control as well as experimental, had the same corrective activities available during a regularly scheduled free time. However, mastery learning students took much greater advantage of these activities. The total time *used* by the experimental group was thus greater than that used by control students, but since the total time *available* was held constant, this was categorized as an equal-time study.

For reasons that were not stated, the implementation of the 15-week chemistry and physics unit was concluded in late January, but the posttests were not given until early June, more than 4 months later. For this reason, the program's effects are listed as retention measures only. Effects favored the mastery learning classes ($ES = +.26$).

Two studies by an Israeli student of Bloom, Zemira Mevarech, produced by far the largest effect sizes of all the mastery learning studies that met the inclusion criteria. One of these (Mevarech, 1980) provided additional time for corrective instruction to the mastery learning classes and is therefore included among the "extra time" studies listed in Table 3. The second (Mevarech, 1986) took place in a "desegregated" Israeli junior high school (i.e., Jews of Middle Eastern and European backgrounds attended the same school). Students were randomly assigned to heterogeneous mastery learning or control classes. The mastery learning classes received lessons and then took formative tests. Students who achieved a criterion score of 70% on biweekly quizzes received corrective instruction from peers or from the teacher, after which a second form of the test was given. Students who achieved the mastery criterion on the first test either served as tutors or worked on enrichment activities. Control classes received the same lessons and materials and were given the formative tests as quizzes, but did not receive corrective instruction.

The outcome measure was an achievement test constructed by the teachers. At the end of the 3-month experiment, mastery learning students scored much higher than control students on this test ($ES = +.90$). The results were also broken down by students' socioeconomic level. Students whose fathers did not complete high school (20% of the sample) gained the most from the mastery learning program ($ES = +1.78$), followed by those whose fathers had a high school degree, 50% of the sample ($ES = +.91$), and those whose fathers completed college ($ES = +.66$).

In light of the extraordinary positive effects found in this study, more than three times the median for all the studies in Table 2, it is useful to consider what may be unique about the Mevarech (1986) study. One important factor is that only two teachers were involved in each treatment, and that while *students* were randomly assigned to treatments, *teachers* were not assigned at random. Thus, the possibility of teacher effects cannot be ruled out; it may be that the mastery learning teachers were simply better teachers than those in the control group. Second, as is the case with all mastery learning studies that used experimenter-made measures, there is a possibility that even though all teachers used the same materials, the mastery learning teachers focused on the specific objectives to be tested more than the control classes did. The posttest was described as covering "various aspects of rational numbers." Because there is much more to a typical algebra course, it may be that the mastery learning classes were spending time mastering a limited set of objectives while the control group may have learned a larger set of objectives (though perhaps at a lower level of mastery).

However, bearing in mind these cautions, the 1986 Mevarech study and the 1980 Mevarech study described later in this article provide some grounds for optimism that certain forms of group-based mastery learning may have strong effects on student achievement.

Overall, the effects summarized in Table 2 could be interpreted as supporting the "curricular focus" claim. The effects of mastery learning on experimenter-made, criterion-referenced measures are generally moderate but consistently positive. Two studies found that the effects of mastery learning were greatest for low achievers, as would be expected from mastery learning theory, and one found effects to be greatest for low-SES students.

However, the meaning of the results summarized in Table 2 is far from clear. The near-zero effects of mastery learning on standardized measures (Table 1) and in particular the dramatically different results for standardized and experimenter-made measures reported by Anderson et al. (1976) suggest that the effects of mastery learning on experimenter-made measures result from a shifting of instructional focus to a particular set of objectives not necessarily more valuable than those pursued by the control group. Unfortunately, it is impossible to determine from reports of mastery learning studies the degree to which control teachers were focusing on the objectives assessed on the experimenter-made measures, yet understanding this is crucial to understanding the effects reported in these studies.

Evidence for the Extra-Time Claim

The problem of unequal time for experimental and control groups is a serious one in mastery learning research in general, but the inclusion criteria used in the present review have the effect of eliminating the studies in which time differences are extreme. Mastery learning studies in which experimental classes receive considerably more instructional time than control classes are always either very brief, rarely more than a week (e.g., Anderson, 1975, 1976; Arlin & Webster, 1983), or they involve individualized or self-paced rather than group-paced instruction (e.g., Jones, 1975; Wentling, 1973). In studies of group-paced instruction conducted over periods of at least 4 weeks, extra time for corrective instruction rarely amounts to more than 20–25% of original time. It might be argued that additional instructional time of this magnitude might be a practicable means of ensuring all students a reasonable level of achievement, and the costs of such an approach might not be far out of line with the costs of current compensatory or special education.

Table 3 summarizes the characteristics and outcomes of group-based mastery learning studies in which the mastery learning classes received extra time for corrective instruction. All four of the studies in this category took place at the secondary level, grades 7–10. Also, these studies are distinctly shorter (5–6 weeks) than were most of the studies listed in Tables 1 and 2.

The median effect size for immediate posttests from the five comparisons in four studies is +.31, but none of three retention measures found significant differences (median *ES* = −.03). However, the four studies differ markedly in experimental procedures, so these medians have little meaning.

The importance of the different approaches taken in different studies is clearly illustrated in a study by Long, Okey, and Yeany (1978). In this study, eighth graders were randomly assigned to six classes, all of which studied the same earth science units on the same schedule. Two classes experienced a mastery learning treatment with teacher-directed remediation. After every two class periods, students in this treatment took a diagnostic progress test. The teacher assigned students specific remedial work, then gave a second progress test. If students still did not achieve at a designated level (the mastery criterion was not described in the article), the teacher tutored them individually. In a second treatment condition, student-directed remediation, students received the same instruction and tests and had the same corrective materials available, but they were asked to use their test results to guide their own learning, rather than having specific activities assigned. These students did not take the second progress test and did not receive tutoring. Students in the third treatment, control, studied the same materials on the same schedule but did not take diagnostic progress tests. Teachers rotated across the three treatments to minimize possible teacher effects.

The results of the Long et al. (1978) study indicated that the teacher-directed remediation (mastery learning) group did achieve considerably more than the control group (*ES* = +.43), but exceeded the student-directed remediation group to a much smaller degree (*ES* = +.19). What this suggests is that simply receiving frequent and immediate feedback on performance may account for a substantial portion of the mastery learning effect. A replication by the same authors (Long et al., 1981) failed to meet the inclusion criteria because it had only one class per treatment. However, it is interesting to note that the replication found the same pattern of effects as the earlier Long et al. (1978) study; the teacher-directed remediation treatment had only slightly more positive effects on student achievement than the student-directed remediation treatment, but both exceeded the control group.

The Long et al. (1978) study included a retention test, which indicated that whatever effects existed at the end of the implementation period had disappeared 12 weeks later. Retention is especially important in studies in which corrective instruction is given outside of class time, as any determination of the cost-effectiveness of additional time should take into account the lasting impact of the expenditure.

Another extra-time study which assessed retention outcomes was a dissertation by Fagan (1976), who randomly assigned four teachers and their 17 seventh-grade classes to mastery learning or control treatments. The mastery learning treatment essentially followed the sequence suggested by Block and Anderson (1975). Students were quizzed at the end of each week, and teachers worked with students who failed to reach an 80% criterion, after which students took a second formative test. The control classes used the same materials and procedures except that they took the formative tests as quizzes. Teachers scored the quizzes, returned them to students, and then went on to the next unit. The teachers followed the same sequence of activities, but were allowed to proceed at their own pace. As a result, the mastery learning classes took 25 days to complete the five units on "transportation and the environment," whereas control classes took only 20–21 days.

Unfortunately, there were pretest differences favoring the control classes of approximately 40% of a grade equivalent on Iowa Test of Basic Skills vocabulary scores. Analyses of covariance on the posttests found no experimental-control differences; in fact, adjusted scores slightly favored the control group ($ES = -.11$). On a 4-week retention measure the control group's advantage was slightly greater ($ES = -.15$). When experimental treatments vary widely in pretests or covariates, statistical adjustments tend to underadjust (see Reichardt, 1979), so these results must be interpreted with caution. However, even discarding the results for the one control teacher whose classes had high pretest scores, differences still favored the control group on the posttest ($ES = -.17$) and on the retention test ($ES = -.23$).

A small study by Hecht (1980) compared mastery learning to control treatments in 10th-grade geometry. Students were randomly assigned to treatments, and each of two teachers taught mastery learning as well as control classes. In the mastery learning classes students were given formative tests every 2 weeks that were followed by "intensive remedial help for those who needed it" (mastery criteria and corrective activities were not stated). Results on an experimenter-made test favored the mastery learning classes ($ES = +.31$).

The largest effect sizes for any of the studies that met the inclusion criteria were found in two studies by Zemira Mevarech. One (Mevarech, 1986) was described earlier. In the second (Mevarech, 1980), students were randomly assigned to eight Algebra I classes in a 2×2 factorial design. One factor was "algorithmic" versus "heuristic" instructional strategies. The "algorithmic" treatments emphasized step-by-step solutions of algebraic problems, focusing on lower cognitive skills. The "heuristic" treatments emphasized problem solving strategies such as Polya's (1957) "understanding-planning-carrying out the plan-evaluating" cycle and focused on higher cognitive skills.

The other factor was mastery learning (feedback-correctives) versus nonmastery. In the mastery learning treatments, students were given formative tests every 2 weeks. They then had three chances to meet the mastery criterion of 80% correct. Corrective instruction included group instruction by the teacher, peer tutoring, and tutoring outside of class time by teachers. The amount of additional time allocated to provide this corrective instruction is not stated, but the author claimed the amount of out-of-class tutoring to be small (Z. Mevarech, personal communication, March 16, 1984). In the nonmastery treatments, students studied the same materials and took the formative tests as quizzes. To hold the different classes to the same schedule, nonmastery classes were given additional problems to work while mastery learning classes were receiving corrective instruction.

The relevant comparisions for the present review involve the mastery learning versus nonmastery factor. Within the algorithmic classes, the mastery learning classes exceeded nonmastery on both "lower mental process" items (i.e., algorithms) ($ES = +.30$) and on "higher mental process" items ($ES = +.77$). Within the heuristic classes, the effects were even greater for both "lower mental process" ($ES = +.66$) and "higher mental process" items ($ES = +.90$).

Overall, the evidence for the "extra time" claim is unclear. Effect sizes for the small number of unequal time studies summarized in Table 3 are no more positive than were those reported for other studies using experimenter-made measures (Table 2), in which mastery learning classes did not receive additional time. In fact, both of the unequal time studies that assessed retention found that any effects observed at posttest disappeared as soon as 4 weeks later. Substantial achievement effects of extra time for corrective instruction appear to depend on provisions of substantial amounts of extra time, well in excess of 20–25%. However, studies in which large amounts of additional time are provided to the mastery learning classes either involved continuous-progress forms of mastery learning or are extremely brief and artificial. What is needed are long-term evaluations of mastery learning models in which corrective instruction is given outside of class time, preferably using standardized measures and/or criterion-referenced measures that register all objectives covered by all classes.

Retention

A total of six comparisons in five studies assessed retention of achievement effects over periods of 4–12 weeks. All six used experimenter-made measures. The median effect size overall is essentially zero, with the largest retention effect ($ES = +.49$) appearing in the Anderson et al. (1976) study, which found no differences on standardized measures.

Discussion

The best evidence from evaluations of practical applications of group-based mastery learning indicates that effects of these methods are moderately positive on experimenter-made achievement measures closely tied to the objectives taught in the mastery learning classes and are essentially nil on standardized achievement measures. These findings may be interpreted as supporting the "weak claim" that mastery learning can be an effective means of holding teachers and students to a specified set of instructional objectives, but do not support the "strong claim" that mastery learning is more effective than traditional instruction given equal time and achievement measures that assess coverage as well as mastery. Further, even this "curricular focus" claim is undermined by uncertainties about the degree to which control teachers were trying to achieve the same objectives as the mastery learning teachers and by a failure to show effects of mastery learning on retention measures.

These conclusions are radically different from those drawn by earlier reviewers and meta-analysts. Not only would a mean effect size across the 17 studies emphasized in this review come nowhere near the mean of around 1.0 claimed by Bloom (1984a, 1984b), Guskey and Gates (1985), Lysakowski and Walberg (1982), or Walberg (1984), but *no single study* reached this level. Only 2 of the 17 studies, both by the same author, had mean effect sizes in excess of the 0.52 mean estimated by Kulik et al. (1986) for precollege studies of mastery testing. How can this gross discrepancy be reconciled?

First, these different reviews focus on very different sets of studies. Almost all of the studies cited in this review would have qualified for inclusion in any of the meta-analyses, but the reverse is not true. For example, of 25 elementary and secondary studies cited by Guskey and Gates (1985), only 6 qualified for inclusion in the present review. Of 19 such studies cited by Kulik et al. (1986), only 4 qualified for inclusion in the present review. Only 2 studies, Lueckemeyer and

Chiappetta (1981) and Slavin and Karweit (1984), appeared in all three syntheses. The list of mastery learning studies synthesized by Lysakowski and Walberg (1982) is short and idiosyncratic, hardly overlapping at all with any of the other reviews, and Bloom's (1984a) article discusses only a few University of Chicago dissertations.

As noted earlier, the principal reason that studies cited elsewhere were excluded in the present paper is that they did not meet the 4-week duration requirement. The rationale for this restriction is that this review focuses on the effects of mastery learning *in practice*, not in theory. It would be difficult to maintain that a 2- or 3-week study could produce information more relevant to classroom practice than a semester- or year-long study, partly because artificial arrangements possible in a brief study could not be maintained over a longer period. Actually, even 4 weeks could be seen as too short a period for external validity. However, it is useful to examine the results of shorter implementations of mastery learning to be sure that arbitrarily drawing a line at 4 weeks' duration does not misrepresent the evidence. A total of 19 elementary and secondary studies with treatment durations of 1 to 3 weeks were cited by Bloom (1984a, 1984b), Kulik et al. (1986), and/or Guskey and Gates (1985). Most of these would not have been excluded from this review on grounds other than their brevity. For example, nine of the studies used self-instructional or programmed materials rather than group-based mastery learning (e.g., Anderson, 1975; Block, 1972), four more used only one class per treatment (e.g. Swanson & Denton, 1977; Tenenbaum, 1982), one used a procedure only tangentially related to mastery learning (Bryant, Fayne, & Gettinger, 1982), and two failed to provide satisfactory evidence that experimental and control classes were initially equivalent (Hymel & Mathews, 1980; Strasler, 1979).

The remaining three studies are all dissertations by Bloom's students. Two of these used very similar procedures. Anania (1983) randomly assigned students in grades 4, 5, and 8 to three treatments: tutoring, mastery learning, and "conventional" instruction. Only the latter two are relevant to the present review. Students in the mastery learning treatment received two 4-day units and one 3-day unit on probability (grades 4–5) or cartography (grade 8). At the end of this time, students took formative tests and then had an extra period in which they received corrective instruction if necessary; if they did not achieve at an 80% criterion level by this time, students might receive additional tutoring after school. Burke (1983) used nearly identical procedures to teach probability to students in grades 4–5. Nordin (1979) compared mastery learning and control methods to teach a 2-week unit on sets to Malaysian sixth graders. In all three studies, mastery learning students far outperformed control; effect sizes were around 1.0 for the Anania and Burke studies, and exceeded 2.0 for the Nordin study.

While all three of these studies are exemplary as basic research, they all have features that severely limit their external validity. First, they all provided significant amounts of extra time for the mastery learning groups, from 125% (Burke, 1983) to 133% (Nordin, 1979) of the time allocated to the control groups. Second, all three selected subject matter that was completely new to students (finessing the issue of student heterogeneity by starting all students at zero) and all three created units that were completely hierarchical, which is to say that learning of the later units depended heavily on mastery of the earlier ones. These procedures are entirely appropriate for theory-building, which was the authors' purpose, but they are hardly representative of conditions in usual classroom teaching, where (for example) students either enter class with different levels of prerequisite skills and/or diverge in their skills over many months or years (see Arlin & Webster, 1983).

In addition to excluding many studies cited elsewhere, the present review included many studies missed in the meta-analyses. These are primarily dissertations and unpublished papers (mostly AERA papers), which comprise 12 of the 17 studies emphasized in this review. Including unpublished studies is critical in any literature review, as they are less likely to suffer from "publication bias," the tendency for studies reporting nonsignificant or negative results not to be submitted to or accepted by journals (see Bangert-Drowns, 1986; Rosenthal, 1979). Other differences in study selection and computation of effect size between the present paper and earlier reviews are important in specific cases. For example, Guskey and Gates (1985) report effect sizes for the Jones, Monsaas, and Katims (1979) study of +.41 for an experimenter-made measure and +.33 for a standardized test, while the present review estimated effect sizes of +.18 and +.09, respectively. The difference is that in the present review pretest differences (in this case favoring the experimental group) were subtracted from the posttest differences. Similarly, Guskey and Gates (1985) report a single effect size of +.58 for the Anderson et al. (1976) study, ignoring the striking difference in effects on standardized as opposed to experimenter-made measures emphasized here.

There are several important theoretical and practical issues raised by the studies of group-based mastery learning reviewed here. These are discussed in the following sections.

Why Are Achievement Effects of Group-Based Mastery Learning So Modest?

The most striking conclusion of the present review is that, other than perhaps focusing teachers and students on a narrow set of objectives, group-based mastery learning has modest to nonexistent effects on student achievement in studies of at least 4 weeks' duration. Given the compelling nature of the theory on which mastery learning is based, it is interesting to speculate on reasons for this.

One possible explanation is that in long-term, practical applications of mastery learning, the quality of training, followup, and/or materials used to support the mastery learning approach are inadequate. One important piece of evidence in support of this possibility comes from a recent study by Dolan and Kellam (1987), who compared an enhanced mastery learning program to the standard mastery learning program used in Baltimore City first grades. The enhanced model provided teachers with 32 hours of instruction in mastery learning principles and practices, monthly progress meetings, classroom visits, files of formative tests, corrective activities, and enrichment activities keyed to school district objectives, and special curriculum materials and other resources to help teachers achieve reading objectives. Teachers using the standard mastery learning procedures also used the teach-test-corrective instruction-test cycle to achieve essentially the same reading objectives, but did not have the additional training, resources, or assistance. A year-long experiment assigned schools and teachers within schools to the two models, and found that the enhanced mastery learning classes gained significantly more on standardized reading tests ($ES = +.39$).

A particular emphasis of the Dolan and Kellam (1987) enhanced mastery learning model was on the quality of the materials used for corrective instruction. In a letter explaining the extraordinary effects obtained in her studies, Zemira

Mevarech (personal communication, January 25, 1987) also emphasized that the quality of the corrective procedures was a key factor in the success of her programs, noting that "corrective activities should be creative, attractive, and designed explicitly to remediate the skills that have not been mastered."

Another possible explanation for the disappointing findings of studies of group-based mastery learning is that it is not only that the *quality* of corrective instruction is lacking, but also that the *amount* of corrective instruction is simply not enough to remediate the learning deficits of low achievers. In none of the studies emphasized in this review did corrective instruction occupy more than one period per week, or 20% of all instructional time. This may be enough to get students up to criterion on very narrowly defined skills, but not enough to identify and remediate serious deficits, particularly when corrective instruction is given in group settings or by peer tutors (as opposed to adult tutors). Studies of students' pace through individualized materials routinely find that the slowest students require 200–600% more time than the fastest students to complete the same amount of material (Arlin & Westbury, 1976; Carroll, 1963; Suppes, 1964), far more than what schools using mastery learning are likely to be able to provide for corrective instruction (Arlin, 1982).

The amount of corrective instruction given in practical applications of group-based mastery learning may be not only too little but also too late. It may be that even 1 or 2 weeks is too long to wait to correct students' learning errors, and many studies provided corrective instruction less frequently, every 3 to 4 weeks. If each day's learning is a prerequisite for the next day's lesson, then perhaps detection and remediation of failures to master individual skills needs to be done daily to be effective. Further, in most applications of mastery learning, students may have years of accumulated learning deficits that 1 day per week of corrective instruction is unlikely to remediate.

Time for corrective instruction in group-based mastery learning is purchased at a cost in terms of slowing instructional pace. If this time does not produce a substantial impact on the achievement of large numbers of students, then a widespread though small negative impact on the learning of the majority may balance a narrow positive impact on the learning of the few students whose learning problems are large enough to need corrective instruction but small enough to be correctable in one class period per week or less.

However, it may be that the feedback-corrective cycle evaluated in the studies reported here is simply insufficient in itself to produce a substantial improvement in student achievement. As Bloom (1980, 1984b) has noted, there are many variables other than feedback-correction that should go into an effective instructional program. Both the process of learning and the process of instruction are so complex that it may be unrealistic to expect large effects on broadly based achievement measures from any one factor; instructional quality, adaptation to individual needs, motivation, and instructional time may all have to be impacted at the same time to produce such effects (see Slavin, 1987).

Is Mastery Learning a Robin Hood Approach to Instruction?

Several critics of mastery learning (e.g., Arlin, 1984a; Resnick, 1977) have wondered whether mastery learning simply shifts a constant amount of learning from high to low achievers. The evidence from the present review is not inconsistent with that view; in several studies positive effects were found for low achievers only. In fact, given that overall achievement means are not greatly improved by group-based mastery learning, the reductions in standard deviations routinely seen in studies of these methods and corresponding decreases in correlations between pretests and posttests are simply statistical indicators of a shift in achievement from high to low achievers. However, it is probably more accurate to say that group-based mastery learning trades coverage for mastery. Because rapid coverage is likely to be of greatest benefit to high achievers, whereas high mastery is of greatest benefit to low achievers, resolving the coverage-mastery dilemma as recommended by mastery learning theorists is likely to produce a "Robin Hood" effect as a by-product.

However, it is important to note that few mastery learning studies have found the method to be *detrimental* to the achievement of high achievers. This may be the case because the coverage versus mastery dilemma exists in *all* whole-class, group-paced instruction, including traditional instruction. For example, Arlin and Westbury (1976) compared individualized instruction to whole-class instruction and found that the instructional pace set by the teachers using the whole-class approach was equal to that of students in the 23rd percentile in the individualized classes, supporting Dahlloff's (1971) contention that teachers set their instructional pace according to the needs of a "steering group" of students in the 10th to 25th percentiles of the class ability distribution. Assuming that an instructional pace appropriate for students at the 23rd percentile is too slow for higher achievers (Barr, 1974, 1975), then whole-class instruction in effect holds back high achievers for the benefit of low achievers. Group-based mastery learning may thus be accentuating a "Robin Hood" tendency already present in the class-paced traditional models to which it has been compared.

The coverage versus mastery dilemma and the corresponding "Robin Hood" effect are problematic only within the context of group-based mastery learning and (at least in theory) only when instruction time is held constant. In continuous-progress or individualized forms of mastery learning in which students can move through material more or less at their own rates, the coverage-mastery dilemma is much less of a concern (Arlin & Westbury, 1976). This does not imply that continuous-progress forms of mastery learning are necessarily more effective than group-based forms; individualization solves the instructional pace problem but creates new problems, such as the difficulty of providing adequate direct instruction to students performing at many levels (Slavin, 1984b). However, there are examples of continuous-progress mastery learning programs that have positive effects on standardized achievement tests (see, e.g., Cohen, 1977; Cohen & Rodriquez, 1980; Slavin & Karweit, 1985; Slavin & Madden, 1987; Slavin, Madden, & Leavey, 1984).

Importance of Frequent, Criterion-Referenced Feedback

Even if we accept the "weak claim" that mastery learning is an effective means of holding teachers and students to a valuable set of instruction objectives, there is still some question as to which elements of mastery learning account for its effects on experimenter-made, criterion-referenced measures. There is some evidence that much of this effect may be accounted for by frequent testing and feedback to students rather than the entire feedback-corrective cycle. Kulik et al. (1986) report that mastery learning studies that failed to control for frequency of testing produced mean effect sizes almost twice those associated with studies in which mastery learning and control classes were tested with equal frequency. Long et al. (1978) compared mastery learning to a condition with the same frequency of testing and found a much smaller effect than in a comparison with a control group that did not receive tests. Looking across other studies, the pattern is complicated by the fact that most that held testing frequency constant also held the control groups to a slower pace than they might otherwise have attained.

Practical Implications

The findings of the present review should not necessarily be interpreted as justifying an abandonment of mastery learning, either as an instructional practice or as a focus of research. Several widely publicized school improvement programs based on mastery learning principles have apparently been successful (e.g., Abrams, 1983; Levine & Stark, 1982; Menahem & Weisman, 1985; Robb, 1985), and many effective nonmastery learning instructional strategies incorporate certain elements of mastery learning—in particular, frequent assessment of student learning of well-specified objectives and basing teaching decisions on the results of these assessments. Further, the idea that students' specific learning deficits should be remediated immediately instead of being allowed to accumulate into large and general deficiencies makes a great deal of sense. It may be that more positive results are obtained in continuous-progress forms of mastery learning, in which students work at their own levels and rates. Use of remedial (e.g., Chapter I), special education, or other resources to provide substantial amounts of instructional time to help lower achieving students keep up with their classmates in critical basic skills may also increase student achievement (Slavin & Madden, 1987). This review concerns only the achievement effects of the group-based form of mastery learning (Block & Anderson, 1975) most commonly used in elementary and secondary schools.

The "2-Sigma Problem" Revisited

One major implication of the present review is that the "2-sigma" challenge proposed by Bloom (1984a, 1984b) is probably unrealistic, certainly within the context of group-based mastery learning. Bloom's claim that mastery learning can improve achievement by more than 1 sigma ($ES = +1.00$) is based on brief, small, artificial studies that provided additional instructional time to the experimental classes. In longer term and larger studies with experimenter-made measures, effects of group-based mastery learning are much closer to ¼ sigma, and in studies with standardized measures there is no indication of any positive effect at all. The 2-sigma challenge (or 1-sigma claim) is misleading out of context and potentially damaging to educational research both within and outside of the mastery learning tradition, as it may lead researchers to belittle true, replicable, and generalizable achievement effects in the more realistic range of 20–50% of an individual-level standard deviation. For example, an educational intervention that produced a reliable gain of .33 each year could, if applied to lower class schools, wipe out the typical achievement gap between lower and middle-class children in 3 years—no small accomplishment. Yet the claims for huge effects made by Bloom and others could lead researchers who find effect sizes of "only" .33 to question the value of their methods.

Clearly, much more research is needed to explore the issues raised in this review. More studies of practical, long-term applications of mastery learning assessing the effects of these programs on broadly based measures of achievement that register coverage as well as mastery are especially needed; idiosyncratic features of the seven studies that used standardized tests preclude any interpretation of those studies as evidence that group-based mastery learning is *not* effective. There is very little known about what would be required to make group-based mastery learning instructionally effective; the Mevarech (1980, 1986) and Dolan and Kellam (1987) studies provide some clues along these lines, but much more needs to be known. In addition, studies carefully examining instructional pace in mastery and non-mastery models are needed to shed light on the coverage-mastery dilemma discussed here. Mastery learning models in which Chapter I or other remedial teachers provide significant amounts of corrective instruction outside of regular class time might be developed and evaluated, as well as models providing daily, brief corrective instruction rather than waiting for learning deficits to accumulate over 1 or more weeks. The disappointing findings of the studies discussed in this review counsel not a retreat from this area of research but rather a redoubling and redirection of efforts to understand how the compelling theories underlying mastery learning can achieve their potential in practical application.

Mastery learning theory and research has made an important contribution to the study of instructional methods. However, to understand this contribution it is critical to fully understand the conditions under which mastery learning has been studied, the measures that have been used, and other study features that bear on the internal and external validity of the findings. This best-evidence synthesis has attempted to clarify what we have learned from research on mastery learning in the hope that this knowledge will enrich further research and development in this important area.

References

Abrams, J. D. (1983). Overcoming racial stereotyping in Red Bank, New Jersey. *School Administrator, 40,* 5, 7.

Anania, J. (1983). The influence of instructional conditions on student learning and achievement. *Evaluation in Education: An International Review Series, 7,* 3–76.

Anderson, L. W. (1975, April). *Time to criterion: An experimental study.* Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Anderson, L. W. (1976). An empirical investigation of individual differences in time to learn. *Journal of Educational Psychology, 68,* 226–233.

Anderson, L. W. (1985). A retrospective and prospective view of Bloom's "learning for mastery." In M. C. Wang & H. J. Walberg (Eds.), *Adapting instruction to individual differences* (pp. 254–268). Berkeley, CA: McCutchan.

Anderson, L. W., Scott, C., & Hutlock, N. (1976, April). *The effects of a mastery learning program on selected cognitive, affective and ecological variables in grades 1 through 6.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Arlin, M. (1973). *Learning rate and learning rate variance under mastery learning conditions.* Unpublished doctoral dissertation, University of Chicago.

Arlin, M. (1982). Teacher responses to student time differences in mastery learning. *American Journal of Education, 90,* 334–352.

Arlin, M. (1984a). Time variability in mastery learning. *American Educational Research Journal, 21,* 103–120.

Arlin, M. (1984b). Time, equality, and mastery learning. *Review of Educational Research, 54,* 65–86.

Arlin, M., & Webster, J. (1983). Time costs of mastery learning. *Journal of Educational Psychology, 75,* 187–195.

Arlin, M., & Westbury, I. (1976). The leveling effect of teacher pacing on science content mastery. *Journal of Research on Science Teaching, 13,* 213–219.

Bangert-Drowns, R. L. (1986). A review of developments in meta-analytic method. *Psychological Bulletin, 99,* 388–399.

Barr, R. C. (1974). Instructional pace differences and their effect on reading acquisition. *Reading Research Quarterly, 9,* 526–554.

Barr, R. C. (1975). How children are taught to read: Grouping and pacing. *School Review, 83,* 479–498.

Block, J. H. (1972). Student learning and the setting of mastery performance standards. *Educational Horizons, 50,* 183–191.

Block, J. H., & Anderson, L. W. (1975). *Mastery learning in classroom instruction.* New York: MacMillan.

Block, J. H., & Burns, R. B. (1976). Mastery learning. In L. S. Shulman (Ed.), *Review of Research in Education* (Vol. 4, pp. 3–49). Itasca, IL: F. E. Peacock.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1,* 2.

Bloom, B. S. (1976). *Human characteristics and school learning.* New York: McGraw-Hill.

Bloom, B. S. (1980). The new direction in educational research: Alterable variables. *Phi Delta Kappan, 62,* 382–385.

Bloom, B. S. (1984a). The 2 sigma problem: The search for methods of instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Bloom, B. S. (1984b). The search for methods of group instruction as effective as one-to-one tutoring. *Educational Leadership, 41* (8), 4–17.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning.* New York: McGraw-Hill.

Bryant, N. D., Fayne, H. R., & Gettinger, M. (1982). Applying the mastery model to sight word instruction for disabled readers. *Journal of Experimental Education, 50,* 116–121.

Burke, A. J. (1983). *Students' potential for learning contrasted under tutorial and group approaches to instruction.* Unpublished doctoral dissertation, University of Chicago.

Cabezon, E. (1984). *The effects of marked changes in student achievement pattern on the students, their teachers, and their parents: The Chilean case.* Unpublished doctoral dissertation, University of Chicago.

Carroll, J. B. (1963). A model for school learning. *Teachers College Record, 64,* 723–733.

Chance, C. E. (1980). *The effects of feedback/corrective procedures on reading achievement and retention.* Unpublished doctoral dissertation, University of Chicago.

Cohen, S. A. (1977). Instructional systems in reading: A report on the effects of a curriculum design based on a systems model. *Reading World, 16,* 158–171.

Cohen, S. A., & Rodriquez, S. (1980). Experimental results that question the Ramirez-Castaneda model for teaching reading to first grade Mexican Americans. *The Reading Teacher, 34,* 12–18.

Collins, K. M. (1971). *An investigation of the variables of Bloom's mastery learning model for teaching junior high school math.* Unpublished doctoral dissertation, Purdue University, Lafayette, IN.

Cooley, W. W., & Leinhardt, G. (1980). The instruction dimensions study. *Educational Evaluation and Policy Analysis, 2,* 7–25.

Crotty, E. K. (1975). An experimental comparison of a mastery learning and lecture-discussion approach to the teaching of world history. *Dissertation Abstracts International, 36,* 7150A. (University Microfilms No. 76-10610)

Dahl, P. R. (1979). An experimental program for teaching high speed word recognition and comprehension skills. In J. E. Button, T. C. Lovitt, & T. D. Rowland (Eds.), *Communication* *research in learning disabilities and mental retardation* (pp. 33–65). Baltimore, MD: University Park Press.

Dahlloff, U. (1971). *Ability grouping, content validity, and curriculum process analysis.* New York: Teachers College Press.

Dillashaw, F. D., & Okey, J. R. (1983). Effects of a modified mastery learning strategy on achievement, attitudes, and on-task behavior of high school chemistry students. *Journal of Research in Science Teaching, 20,* 203–211.

Dolan, L. J., & Kellam, S. G. (1987). *Preliminary report of mastery learning intervention during first grade with cohort I students.* Baltimore, MD: Johns Hopkins University, Department of Mental Hygiene, Prevention Center.

Dunkelberger, G. E., & Heikkinen, H. (1984). The influence of repeatable testing on retention in mastery learning. *School Science and Mathematics, 84,* 590–597.

Fagan, J. S. (1976). Mastery learning: The relationship of mastery procedures and aptitude to the achievement and retention of transportation-environmental concepts by seventh grade students. *Dissertation Abstracts International, 36,* 5981. (University Microfilms No. 76-6402)

Fitzpatrick, K. A. (1985, April). *Group-based mastery learning: A Robin Hood approach to instruction?* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Froemel, J. (1980). *Cognitive entry behaviors, instructional conditions, and achievement: A study of their interrelationships.* Unpublished doctoral dissertation, University of Chicago.

Fuchs, L. S., Tindal, F., & Fuchs, D. (1985). *A comparison of mastery learning procedures among high and low ability students.* Unpublished manuscript. Vanderbilt University, Nashville, TN. (ERIC Document Reproduction Service No. ED 259 307)

Gettinger, M. (1985). Time allocated and time spent relative to time needed for learning as determinants of achievement. *Journal of Educational Psychology, 77,* 3–11.

Glass, G., Cohen, L., Smith, M. L., & Filby, N. (1982). *School class size.* Beverly Hills, CA: Sage.

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

Guskey, T. R. (1982). The effects of change in instructional effectiveness on the relationship of teacher expectations and student achievement. *Journal of Educational Research, 75,* 345–349.

Guskey, T. R. (1984). The influence of change in instructional effectiveness upon the affective characteristics of teachers. *American Educational Research Journal, 21,* 245–259.

Guskey, T. R., & Gates, S. L. (1985, April). *A synthesis of research on group-based mastery learning programs.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Guskey, T. R., & Gates, S. L. (1986). Synthesis of research on the effects of mastery learning in elementary and secondary classrooms. *Educational Leadership, 43* (8), 73–80.

Gutkin, J. (1985, March). *The effect of diagnostic inservice training on class reading achievement and the number of lessons covered.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Hecht, L. W. (1980, April). *Stalking mastery learning in its natural habitat.* Paper presented at the annual meeting of the American Educational Research Association, Boston.

Hymel, G. M. (1982). *Mastery learning: A comprehensive bibliography.* New Orleans: Loyola University.

Hymel, G. M., & Mathews, G. (1980). Effects of a mastery approach on social studies achievement and unit evaluation. *Southern Journal of Education Research, 14,* 191–204.

Jones, B. F., Monsaas, J. A., & Katims, M. (1979, April). *Improving reading comprehension: Embedding diverse learning strategies within a mastery learning instructional format.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Jones, F. G. (1975). The effects of mastery and aptitude on learning, retention, and time. *Dissertation Abstracts International, 35,* 6547. (University Microfilms No. 75-8126)

Katims, M., & Jones, B. F. (1985). Chicago Mastery Learning Reading: Mastery learning instruction and assessment of inner-city schools. *Journal of Negro Education, 54,* 369–387.

Katims, M., Smith, J. K., Steele, C., & Wick, J. W. (1977, April). *The Chicago Mastery Learning Reading Program: An interim evaluation.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Keller, F. S. (1968). "Good-bye, teacher . . ." *Journal of Applied Behavior Analysis, 1,* 78–89.

Kersh, M. E. (1970). *A strategy for mastery learning in fifth grade arithmetic.* Unpublished doctoral dissertation, University of Chicago.

Kulik, C. L., Kulik, J. A., & Bangert-Drowns, R. L. (1986, April). *Effects of testing for mastery on student learning.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Kulik, J. A., Kulik, C. L., & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's Personalized System of Instruction. *American Psychologist, 34,* 307–318.

Levine, D. U. (Ed.). (1985). *Improving student achievement through mastery learning programs.* San Francisco: Jossey-Bass.

Levine, D. U., & Stark, J. (1982). Instructional and organizational arrangements that improve achievement in inner-city schools. *Educational Leadership, 39,* 41–46.

Leyton, F. S. (1983). *The extent to which group instruction supplemented by mastery of initial cognitive prerequisites approximates the learning effectiveness of one-to-one tutorial methods.* Unpublished doctoral dissertation, University of Chicago.

Long, J. C., Okey, J. R., & Yeany, R. H. (1978). The effects of diagnosis with teacher- or student-directed remediation on science achievement and attitudes. *Journal of Research in Science Teaching, 15,* 505–511.

Long, J. C., Okey, J. R., & Yeany, R. H. (1981). The effects of a diagnostic-prescriptive teaching strategy on student achievement and attitude in biology. *Journal of Research in Science Teaching, 18,* 515–523.

Lueckemeyer, C. L., & Chiappetta, E. L. (1981). An investigation into the effects of a modified mastery learning strategy on achievement in a high school human physiology unit. *Journal of Research in Science Teaching, 18,* 269–273.

Lysakowski, R., & Walberg, H. (1982). Instructional effects of cues, participation, and corrective feedback: A quantitative synthesis. *American Educational Research Journal, 19,* 559–578.

Mathews, G. S. (1982). *Effects of a mastery learning strategy on the cognitive knowledge and unit evaluation of students in high school social studies.* Unpublished doctoral dissertation, University of Southern Mississippi, Hattiesburg.

Menahem, M., & Weisman, L. (1985). Improving reading ability through a mastery learning program: A case study. In D. Levine (Ed.), *Improving student achievement through mastery learning programs* (pp. 223–240). San Francisco: Jossey-Bass.

Mevarech, Z. R. (1980). *The role of teaching-learning strategies and feedback-corrective procedures in developing higher cognitive achievement.* Unpublished doctoral dissertation, University of Chicago.

Mevarech, Z. R. (1985a). The effects of cooperative mastery learning strategies on mathematical achievement. *Journal of Educational Research, 78,* 372–377.

Mevarech, Z. R. (1985b, April). *Cooperative mastery learning strategies.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Mevarech, Z. R. (1986). The role of a feedback corrective procedure in developing mathematics achievement and self-concept in desegregated classrooms. *Studies in Educational Evaluation, 12,* 197–203.

Mueller, D. J. (1976). Mastery learning: Partly boon, partly boondoggle. *Teachers College Record, 78,* 41–52.

Nordin, A. B. (1979). *The effects of different qualities of instruction on selected cognitive, affective, and time variables.* Unpublished doctoral dissertation, University of Chicago.

Okey, J. R. (1974). Altering teacher and pupil behavior with mastery teaching. *School Science and Mathematics, 74,* 530–535.

Okey, J. R. (1977). The consequences of training teachers to use a mastery learning strategy. *Journal of Teacher Education, 28* (5), 57–62.

Polya, G. (1957). *How to solve it* (2nd ed.). New York: Doubleday.

Rankin, P. T., Anderson, C. T., & Bergman, W. G. (1936). Ability grouping in the Detroit individualization experiment. In Whipple, G. M. (Ed.), *The grading of pupils (35th yearbook of the National Society for the Study of Education, Part I,* pp. 277–288). Bloomington, IL: Public School Publishing Co.

Reichardt, C. S. (1979). The statistical analysis of data from nonequivalent group designs. In T. C. Cook & D. T. Campbell, *Quasi-experimentation: Design and analysis issues for field settings* (pp. 147–205). Chicago: Rand McNally.

Resnick, L. B. (1977). Assuming that everyone can learn everything, will some learn less? *School Review, 85,* 445–452.

Robb, D. W. (1985). Strategies for implementing successful mastery learning programs: Case studies. In D. Levine (Ed.), *Improving student achievement through mastery learning programs* (pp. 255–272). San Francisco: Jossey-Bass.

Rosenthal, R. (1979). The "file-drawer problem" and tolerance for null results. *Psychological Bulletin, 92,* 500–504.

Rubovits, J. J. (1975). A classroom field experiment in mastery learning. *Dissertation Abstracts International, 36,* 2720-A. (University Microfilms No. 75-24395)

Slavin, R. E. (1983). *Cooperative learning.* New York: Longman.

Slavin, R. E. (1984a). Meta-analysis in education: How has it been used? *Educational Researcher, 13* (8), 6–15, 24–27.

Slavin, R. E. (1984b). Component building: A strategy for research-based instructional improvement. *Elementary School Journal, 84,* 255–269.

Slavin, R. E. (1986a). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15* (9), 5–11.

Slavin, R. E. (1986b). *Ability grouping and student achievement in elementary schools: A best-evidence synthesis* (Tech. Rep. No. 1). Baltimore, MD: Johns Hopkins University, Center for Research on Elementary and Middle Schools.

Slavin, R. E. (1987). A theory of school and classroom organization. *Educational Psychologist, 22,* 89–108.

Slavin, R. E., & Karweit, N. L. (1984). Mastery learning and student teams: A factorial experiment in urban general mathematics classes. *American Educational Research Journal, 21,* 725–736.

Slavin, R. E., & Karweit, N. L. (1985). Effects of whole-class, ability grouped, and individualized instruction on mathematics achievement. *American Educational Research Journal, 22,* 351–367.

Slavin, R. E., & Madden, N. A. (1987, April). *Effective classroom programs for students at risk.* Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Slavin, R. E., Madden, N. A., & Leavey, M. (1984). Effects of Team Assisted Individualization on the mathematics achievement of academically handicapped and nonhandicapped students. *Journal of Educational Psychology, 76,* 813–819.

Smith, J. K. (1981, April). *Philosophical considerations of mastery learning theory.* Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

Strasler, G. M. (1979, April). *The process of transfer in a learning for mastery setting.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Strasler, G. M., & Rochester, M. (1982, April). *A two-year evaluation of a competency-based secondary school project in a learning for mastery setting.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Suppes, P. (1964). Modern learning theory and the elementary school curriculum. *American Educational Research Journal, 2,* 79–93.

Swanson, D. H., & Denton, J. J. (1977). Learning for mastery versus personalized system of instruction: A comparison of remediation strategies with secondary school chemistry students. *Journal of Research in Science Teaching, 14,* 515–524.

Taylor, B. R. (1973). *The effect of a mastery learning of minimum essentials requirement on achievement in ninth grade algebra.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Tenenbaum, G. (1982). *A method of group instruction which is as effective as one-to-one tutorial instruction.* Unpublished doctoral dissertation, University of Chicago.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41* (8), 19–27.

Wentling, T. L. (1973). Mastery versus nonmastery instruction with varying test item feedback treatments. *Journal of Educational Psychology, 65,* 50–58.

Wortham, S. C. (1980). *Mastery learning in secondary schools: A first year report.* Unpublished manuscript, San Antonio (TX) Public Schools. (ERIC Document Reproduction Service No. ED 187 387)

Wyckoff, D. B. (1974). *A study of mastery learning and its effects on achievement of sixth grade social studies students.* Unpublished doctoral dissertation, Georgia State University, Atlanta.

Author

ROBERT E. SLAVIN, Director, Elementary School Program, Center for Research on Elementary and Middle Schools, Johns Hopkins University, Baltimore, MD 21218. *Specializations*: cooperative learning, school and classroom organization, desegregation, mainstreaming, field research methods.