
1

Understanding Mastery Learning

The movement toward higher academic standards for our children carries with it a number of implications, some explicit, some implied:

- That there is some consensus on our academic goals and priorities
- That those goals are both explicable and measurable
- That we can validly differentiate between truly competent, incompetent, or incomplete achievement in those academic domains
- That we have the resources and the will to remediate and motivate those who are initially unprepared for, or slow to grasp, one or more of those academic goals
- That in attaining those new higher standards, we model fair and equitable procedures

It is our thesis that all the above points are intimately connected to mastery learning (and criterion-referenced assessment). Indeed, the standards for each of the professional disciplines are amply stocked with phrases such as “All children can learn,” “Aim for mastery,” and “Students should be able to demonstrate competent levels of achievement.”

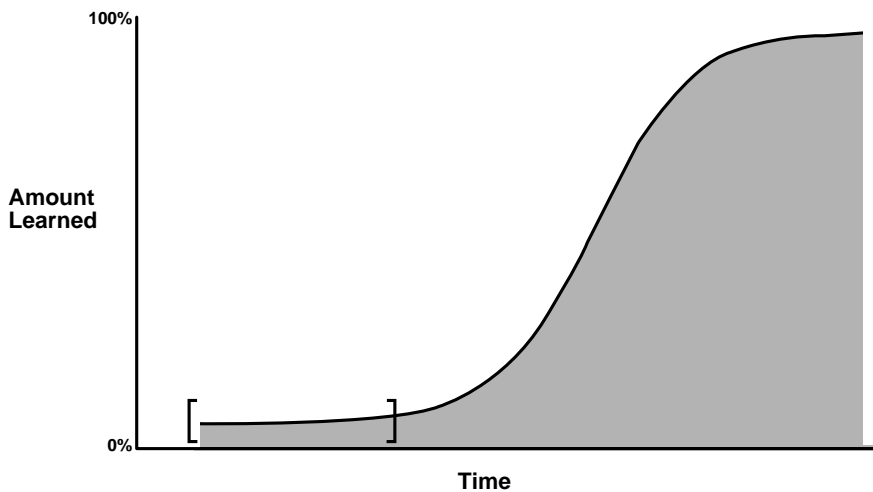
The problem is that although the connection between teaching to standards and mastery learning is logical and natural, only a small proportion of teachers and schools have adopted mastery learning. Moreover, many of those who claim to be using mastery learning make one or more of the following common, often fatal, mistakes: (a) Passing a mastery test is conceptualized as the endpoint instead of the initial stage of the learning/memory process; (b) there is no requirement and concomitant grading incentive to go beyond initial mastery; (c) mastery testing is embedded in an overall grading scheme (often a leftover norm-referenced and competitive scheme) that contradicts the goal of achieving mastery by all (a criterion-referenced purpose); (d) demonstrations of mastery are limited to objective tests at the knowledge/recall end of the thinking continuum (e.g., Bloom's 1956 taxonomy); and therefore (e) students are overtested and underchallenged.

The above errors are common, we believe, because educators tend to think of mastery learning as a teaching technique or a testing procedure, instead of a philosophy or theory that provides the basis for decisions about techniques or assessments. Furthermore, they do not understand the theoretical and empirical bases on which mastery learning is founded.

It is our purpose in this introductory chapter to briefly describe each of the following foundations for mastery learning: the learning/memory base, the measurement base, the theoretical bases (including competency vs. helplessness, Erikson's social development theory, and Carroll's model), the brain base, and the empirical base. The accumulated weight of these positions will lead us to some conclusions about teaching, assessment, and grading. These, in turn, will be directly relevant to the standards movement of our time and why these new standards will fail if we do not take heed of the lessons from mastery learning.

The Learning/Memory Base

Consider a learning task: memorizing multiplication facts or vocabulary; acquiring a skill, such as playing a piece on the piano or hitting a tennis ball with topspin; comprehending a paragraph or theory; or learning a new teaching technique or thinking strategy. Each of these, from the simplest to the most complex, has a number of essential prerequisites—that is, some prior knowledge that if previously acquired and currently activated can facilitate achievement of

Figure 1.1 The S-Shaped Learning Curve

the task at hand. For multiplication facts, two likely essentials are one-to-one correspondence for rational counting, and skill with addition. For thinking strategies, metacognitive skills (such as the ability to actively monitor what you comprehend and what you don't) and the ability to organize and categorize are likely prerequisites. If such prior knowledge or skill is absent, then the learning task is more difficult. And to complete the scenario, if the prerequisite knowledge has been mislearned or is otherwise inaccurate, acquisition of the new task is further confounded with the need to unlearn prior misconceptions.

Teachers acknowledge the differences in prior learning among their charges in regard to their *readiness*. We speak of Mary, who has great facility for foreign languages, or Bill, who is not yet ready to read. In the terminology of the famous S-shaped learning curve (see Figure 1.1), readiness is manifested in how long the line is before any progress or acceleration can be observed. Students like Mary, who have mastered prerequisites, may show immediate progress, whereas students like Bill may need hours or days trying to figure out what the task is or to acquire the prerequisites, and thus remain in the bracketed part of the curve for an extended period.

Learning, in other words, occurs in phases or episodes, and this *original learning phase* includes (a) the readiness component (described above), (b) learning to initial mastery, and (c) forgetting. Although forgetting has not been mentioned up to now, it is clear that forgetting is the inevitable result of initial learning, even when a high mastery standard of, say, 80% to 100% correct is required. When the degree of

original learning is less than mastery, say, 60% to 80%, then forgetting is likely to occur more rapidly or be more complete. If it is less than 60%, it is questionable to speak of forgetting at all, because learning was inadequate in the first place.¹

Students show that they understand this principle implicitly when they ask, “Why do we have to learn this stuff anyway? We’ll only forget it.” Our typical answers, “Because it will be on the test” or “Because I said so,” are not satisfactory. In fact, we have been able to find only one satisfactory answer to the question, and it was supplied in one of the first empirical studies of learning/forgetting (Ebbinghaus, 1885/1964). The answer is that relearning is faster—that is, there is a considerable *savings* of time in relearning compared with original learning. Furthermore, there is a positive relationship between amount of time saved in relearning and the degree of original learning, with essentially no savings when original learning is below some acceptable threshold (which we earlier argued was 60% or less).

If original acquisition and forgetting constitute Phase 1 of learning, then each new relearning and forgetting episode constitutes an additional phase. In addition, for material or skills that were mastered in Phase 1, each new relearning episode constitutes additional amounts of *overlearning*, defined as practice beyond initial learning, which is inversely related to forgetting.

Each new relearning-forgetting phase also provides the opportunity for *distributed practice*—that is, rehearsal of the material over hours, days, and weeks. When compared with *massed practice* (i.e., cramming, or including the same number of learning episodes into one session), the positive effects of distributed practice on memory are widely known, as advertised by teachers and coaches when they say, “Practice a little each day.” From a cognitive constructivist point of view, distributed practice also provides experience retrieving previously stored material, comparing new examples with those stored, reorganizing what is known, and recoding it for memory storage and subsequent accessibility.

Although the above material has been widely known for decades and documented in many basic learning as well as educational psychology texts,² it is not usually considered in the context of mastery learning. And yet the implications are straightforward and powerful, as both a positive and a negative example will portray.

For the positive example, consider the experience we teachers have. The first year of teaching a unit, we have a lot to learn (even though many prerequisites have already been mastered in college and student teaching). The next year, when we reach that unit again, we find we’ve forgotten quite a bit. Fortunately, relearning is faster,

and we find ourselves reorganizing the material, coming up with new examples, and so forth. The next year, forgetting has been less yet, and thus there is greater savings in relearning. By the 10th year, the material is almost totally recalled, with examples virtually falling off the tongue. The material seems so easy by this time that many teachers can now be heard complaining, "The students are getting dumber and dumber every year." Sadly, this is one of the negative effects of becoming expert in something: We lose empathy for the novice. (Note the parallel to what happens once a nonconservator on Piaget's tasks becomes a conservator: e.g., the 8-year-old who understands that the amount of water does not change when poured into a taller, thinner glass cannot recall that she expected more in the taller, thinner glass when she was 5). This is also what distinguishes a mere expert from a teacher: An expert can do it; a teacher can do it but also remembers what it takes to progress from novice to expert.

For the negative example of how these learning/memory concepts form the basis for mastery learning, consider some students whose original learning is 50% or less. Their forgetting in Phase 1 will be at least as rapid as it will for those who mastered the material and, because they learned so little, more complete. Furthermore, there will be little or no savings in relearning during Phase 2. Then, if this material is treated as a review ("This was covered last year"), less time will be spent on relearning (and after all, students who mastered originally will need less time). Thus the relearning episode for those who need it most will also be substandard, leading to relatively little residue in memory and therefore little or no savings for Phase 3 relearning. By Phase 3, the motivation to learn this material will also be eroding ("I was never very good at this"), an issue we shall explore in more detail in "Learned Helplessness."

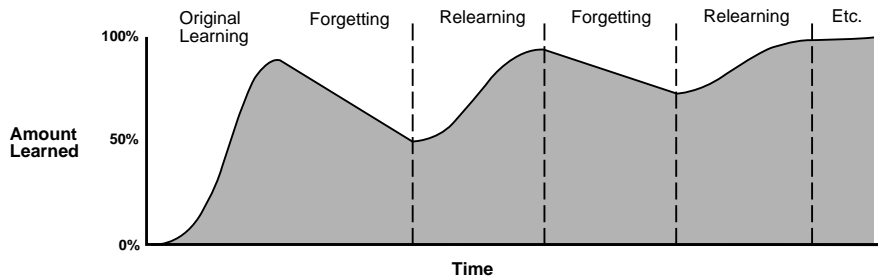
A summary of the long-term effects of the difference between mastery and nonmastery at original learning is provided in Figure 1.2. After four or five episodes, when learners in Part A say to those in Part B, "I forgot more than you ever learned," the sad fact will be that they are telling the truth. A close look at the curve in Part B shows that after four or five episodes, those persons—they can hardly be called learners—have learning/forgetting curves that resemble the brain waves of comatose patients.

The Measurement Base

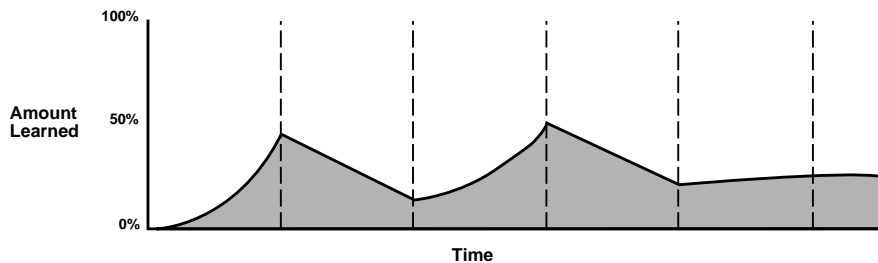
Since at least Glaser's (1963; Glaser & Nitko, 1971) publications, the field of educational measurement has acknowledged the importance

Figure 1.2 Hypothetical Learning/Forgetting Curves for Mastery Versus Nonmastery

A. When Original Learning is Adequate (Mastery)



B. When Original Learning is Inadequate (Nonmastery)



of the distinction between *norm-referenced* (NR) and *criterion-referenced* (CR) assessment. Whereas NR assessment is used to refer to the traditional psychometric approach of measuring and comparing individual differences in relation to norms provided by others, CR assessment interprets a student's score in relation to the goals and criteria of instruction without regard to other students' scores.

NR assessment, therefore, is useful for obtaining rank in class, for estimating a person's status in relation to some normative group, and for selecting the most qualified people from a long list of hopefuls (for awards, for graduate school, etc.). CR assessment, in contrast, is useful for measuring how skillful a person is in relation to a particular instructional domain, what growth a person has demonstrated over time, and what else the person needs to attain specified instructional goals. In Bloom's (1976) succinct contrast, it is a difference between selecting talent and developing talent.

For NR purposes, students take a test or perform a skill and are scored relative to one another or to appropriate norms. The absolute score is not important and is usually not even meaningful, for at least the following reasons. First, the domain being tested is not usually clearly specified and does not have to be. To rank people on IQ or math aptitude, for example, it is best to write items that sample from many areas, not just vocabulary or exponents. Second, to maximize reliability of measurement, each item selected for the test should be missed by about 50% of the examinees; this maximizes the range of the test scores and creates a normal curve. Scores on the test, therefore, are not interpretable, except in relation to the norms; for example, hearing that your child scored 430 on the SAT verbal test tells you nothing about what questions he or she could answer, and indeed, this information needs to be accompanied by statistics such as percentile rank (e.g., 38th percentile) to have any meaning.

Finally, to demonstrate reliability or predictive validity of an NR test, one must show that people stay in approximately the same rank order year to year or test to test. This is independent of absolute score, because the entire population can grow in height, knowledge, or skill, as they do in physical stature, and still the individuals would stay in approximately the same position relative to each other. A prime example of that is, of course, that the top third of readers in first grade are likely also to remain in the top third by fourth grade, whether the whole school improves or not in those 3 years.

For CR purposes, the contrasts with the above are striking. The individual's score *is* important in and of itself because it is intended to be a measure of level of knowledge or skill. First, it makes sense to speak of being a novice or an expert only in some domain, so the domain of the skills must be specified explicitly. Test items or performances must therefore be congruent with instructional goals, making those assessments domain referenced (Hively, 1974; Nitko, 1980; Popham, 1978).

Second, for reliability and validity of measurement, we do not want items to fool half the people all the time. Perhaps we want all the students to fail the test before instruction and all of them to pass after instruction—a measure called “instructional sensitivity” (Haladyna & Roid, 1981), but that's the closest we would want items with difficulty of .5. Because effective instruction should help—indeed, require—students to be achieving at 80% correct or higher on tests, then typical NR measures of reliability and validity will be underestimates. A psychometrically appropriate CR assessment is completed via a standard decision table, such as Table 1.1. This is

Table 1.1 Decisions About Mastery

<i>Level of Competence</i>	<i>Test Results</i>	
	<i>Mastered</i>	<i>Did Not Master</i>
Truly competent	Correct-positive	False-negative
Not competent	False-positive	Correct-negative

because decisions about mastery, like medical diagnoses, fall into one of the following categories (Gentile, 1997, p. 490):

1. A truly competent individual passes the test (a correct-positive decision).
2. A truly competent individual fails the test (a false-negative decision).
3. A truly incompetent individual fails the test (a correct-negative decision).
4. A truly incompetent individual passes the test (a false-positive decision).

This kind of decision table allows the passing standard to be placed wherever it needs to be for the current CR purposes. For example, it might be 70% for initial mastery of multiplication tables, and 85% 2 months later. Changing the passing standard also changes the probability of types of errors: That is, if the standard is too low, we will make more false-positive decisions. How high the standard should be, therefore, depends on how high the stakes are (e.g., Millman, 1989). Given all this and in contrast with NR measures, once these standards are established, each individual's score is directly interpretable as a measure of competence in the domain being tested.

Finally, there are statistical procedures for assessing the reliability and validity of CR assessments, following the logic of Table 1.1 and in parallel with methods of assessing the adequacy of interrater agreements, providing measures such as the percentage of agreements (e.g., Gentile, 1997; Martuza, 1977).

In sum, measurement decisions are fundamentally different for NR and CR purposes, and even if they are correlated (because students who hold the top ranks in class are most likely to pass any given test), it is psychologically different for both teachers and students to conceive of passing a test as a self-actualization rather than as a competition.

Theoretical Bases

Learned Helplessness

Let's return for a moment to the students in Figure 1.2B. If their first experience in learning fractions, say, is unsuccessful, they will forget all or most of it. Next year, they may even claim that they "never had this stuff before," and they will probably believe it. On the second time through, if they are still unsuccessful (which is likely, because teachers usually spend less time reviewing previous material than was spent on initial learning), they will demonstrate little savings and forget again. By the third and fourth exposures to the material, they at least remember they had it before, but may stop trying and explain their lack of motivation with statements such as "I was never very good at math" or "Why do we have to learn this stuff anyway? I'm never going to use it." *Learned helplessness* has set in.

Under laboratory conditions, learned helplessness is developed by first exposing animals or humans to a series of experiences in which failure is inevitable and beyond their control (e.g., Peterson, Maier, & Seligman, 1993; Seligman, 1975). Later, when success is now possible and personally controllable, the victim does not even try. On the emotional level, there is a heightened state of fear, which if prolonged, can easily turn into apathy or depression. On the behavioral/motivational level, there is no perseverance or willingness even for trial-and-error searches (because "Nothing I do ever satisfies these people"). On the cognitive level, there is no discovery of what works, no understanding or organization of an information base, and a long list of defensive excuses or causal attributions, such as "I was never very good at this" and "I could do it if I want to, but school sucks" (the former a primarily female attribution, the latter male; e.g., Dweck & Licht, 1980). It is also more self-protective to adopt a strategy of not trying—or pretending not to try—than to try and not succeed.

Math seems to be a field particularly vulnerable to learned helplessness, because new topics and courses seem to be quite different from previous ones (from multiplication of whole numbers to fractions, arithmetic to algebra to trigonometry, etc.). Even having great success at earlier levels does not immunize against having difficulty on a new topic. Thus even being a "good" student or having 100% success does not guarantee against learned helplessness later, particularly if what students have been good at is memorizing without understanding.

But those primarily at risk for learned helplessness are those who come to school and have not mastered fundamentals (as mentioned

earlier). If we teachers cannot diagnose their problems correctly—and early—they are almost destined to fail to master the new tasks. Sadly, to continue the math example, many teachers are not skilled enough themselves to diagnose a child's problems with addition and subtraction as being a deficit in rational counting or one-to-one correspondence. Thus these students comprise the population in Figure 1.2B.

Is there a cure? As in health, prevention is easier than cure. With his learned-helpless dogs, Seligman (1975) literally had to drag them across the barrier to escape electric shock, anywhere from 25 to 200 times, before they once again tried to explore and control their environment. With humans, whose patterns of thought ("I was never very good at this") may reinforce the helpless-behavior patterns, dragging is more figurative than literal. In any case, the cure for helplessness is competence, and only when students are succeeding do feelings of self-efficacy, self-control, and self-esteem begin to follow (see also Bandura, 1977, 1986).

Psychosocial Development

Virtually every theory of human development speaks of the passages from one stage to another as milestones to be achieved, crises to be resolved, or self-worth to be earned. None treat development as a maturational unfolding. Piaget and Inhelder (1969) suggest that children are motivated to move from preconservation to conservation (e.g., that the amount of liquid does not change when poured into a taller, thinner glass) only when they experience a *cognitive conflict* (e.g., their prediction that there is more water in the taller glass is disconfirmed by pouring it back). Vygotsky's (1962, 1978) writings are laced with terms such as *turning point*, *struggle*, *leap*, and *rupture*. And central to Erikson's theory (1963, 1968), is the concept of *identity crisis*—a different one for each of his eight stages from infancy through old age.

Of most relevance to the present discussion is Erikson's analysis of the developmental tasks of school age, corresponding roughly from age 6 to puberty. The identity crisis to be resolved is that of *industry* versus *inferiority*, in which children are beginning to formulate their identities by what they can do: "I am what I can learn to make work" (Erikson, 1968, p. 127) or "I am what I learn" (Erikson, 1980, p. 87).

During this age range, the developmental tasks set by our technological society concern verbal and numerical literacy, information-handling skills, and the interpersonal abilities to negotiate these in

complex social situations, such as schools. Failure to master these skills results in inferiority, guilt, and lack of self-worth and self-esteem. No amount of well-intended praise for trying—and certainly no amount of providing labels as excuses, such as a specific learning disability—can rectify or compensate for failure at these tasks. Self-worth at this age, readiness for the crises of adolescence, and dreams for the future as an adult—in a word, *identity*—can be achieved only in the old-fashioned way: by *earning* it via competence in these elementary school tasks.

When today's educators say in the various standards that "All children can learn" or in politics that "No child will be left behind," they are recognizing the importance of success at each level of schooling as being preparatory for the next level and for an adult career. We also know, from Erikson's work, that success is even more important than that: It goes to the core of each child's identity. And so we must say to educators, as we say to our children, "If at first you don't succeed, try, try again."

Carroll's Model of School Learning

There are two basic systems of mastery learning (introduced here and discussed in more detail in the Appendix), each derived from different theories but having objectives, a mastery standard, and CR testing in common. The first, Fred Keller's (1968; Keller & Sherman, 1974) Personalized System of Instruction (PSI), is an individually paced system with few large-group lessons. It was developed on a behavioral model in which progress through a curriculum is contingent on successful completion of required and optional assignments (tests, performances, papers, etc.), but this is accomplished by students at their own rates.

The second model, Learning for Mastery (LFM), is a mostly group-based approach, with individualization provided as needed. Like Keller's PSI, it was also published in 1968, by Benjamin Bloom, who derived it from John Carroll's (1963) Model of School Learning. Carroll's model (see also Carroll, 1989) can be succinctly formulated in the following ratio:

$$\text{Amount of Learning} = \frac{\text{Time Actually Spent}}{\text{Time Needed}}$$

That is, what students learn is a function of whether they spend the time they need to learn it.

Contributing to *time spent* are two factors: *opportunity*, defined as the time allowed or scheduled for learning by the teacher, and *perseverance*, which is the time a student is willing or motivated to spend. Contributing to *time needed* are three factors: *quality of instruction*, which includes how well the material is sequenced, presented, and adapted to the learners; *ability to understand instruction*, defined as the extent to which students can comprehend the language of instruction and requirements of the task; and *aptitude*, expressed simply as the time required by an individual to learn some material or skill to some preestablished level.

By Carroll's model, then, we might estimate that Peter and Alicia were likely to need about 3 hours and 1 hour, respectively, to learn how to multiply fractions, based mostly on our experience of teaching them addition and subtraction of fractions by similar methods. If our schedule provides only 2 hours of instruction and practice, however, Peter's learning cannot exceed $\frac{2}{3}$ (two thirds of it learned well or more likely, most of it partially learned), while Alicia can attain mastery (and go beyond $\frac{2}{1}$). If they both miss 1 of the 2 hours of class (by being physically or cognitively absent), then Alicia may still be able to master the concepts ($\frac{1}{1}$), but Peter will have barely begun ($\frac{1}{3}$).

Although it may often appear difficult to quantify these variables, Carroll's model has become far more valuable as a philosophy than as a numerical formula. *Aptitude* had traditionally been defined as an intellectual trait, potential, or capacity; thus Carroll's definition as time needed was radical. It suggested to Bloom (1971, 1981), for example, that for most objectives and standards adopted for schools, all students can learn but they will differ in rates of mastery according to their aptitudes. It furthermore allowed vastly different amounts of time needed by each individual in different domains (e.g., Alicia may have higher aptitude for math than for music). This notion has recently gained further credibility with the popularity of Gardner's (1983, 1993) theory of multiple intelligences.

Whether or not it relates well to other theories, Carroll's view of aptitude has become integral to mastery learning systems, as well as to standards. From phrases such as "All children can learn" to the newly adopted Regents diploma to be required for every high school graduate in New York State, this view is central.

The Brain Base

It is beyond the scope of this book to go into much detail on what has come to be known as "brain-based education," but the popularity of

the movement during the last decade requires at least a few comments. Structurally, the brain is usually described as being comprised of the following major parts (e.g., Sousa, 2001):

1. The *brain stem* (also known as the “reptilian brain”), controls the bodily functions that keep us alive, such as heart rate, respiration, digestion, and attention to signals in the environment that may constitute threats.
2. The *limbic system* (also known as “the old mammalian brain”) generates emotions and consolidates experiences into memories, thus providing continual interplay between cognitive and emotional processes.
3. The *cerebrum*, or cerebral cortex, is divided into two hemispheres (the so-called right and left brains), which are responsible for speech, thinking, and almost every other kind of act we call cognitive.
4. The *cerebellum* coordinates movements and, likely, their relation to sensations, cognitions, and emotions.

Each of these parts comprises other structures, some of which have unique functions but are often capable of duplicating functions from other parts of the brain. Within each structure are billions of cells, the most central of which (for our purposes) are neurons, or nerve cells, which function to make connections with other neurons. They do this by extending dendrites, or branches, to other neurons and making synapses, or points of connection across which electrical and chemical impulses travel; this shows up as activity in brain-imaging research and diagnosis.

There is much we don't know about these processes, but two apparently supportable generalizations do seem relevant to the current discussion:

1. *Use it or lose it:* Synaptic connections that are repeatedly stimulated proliferate, while others are eliminated or pruned (e.g., Bruer, 1997).
2. *The brain works as a whole:* Contrary to simplistic right brain-left brain conceptions, environmental stimulation activates many parts of the brain, depending on which patterns (in #1) have been established.

In perhaps the first theory of how the brain learns, Donald Hebb (1959) suggested that repeatedly stimulated connections produced *cell assemblies*, which work as neuronal teams to allow for automatic and relatively effortless performances when an important skill is mastered. This occurs, as noted above, not just because of the connections that are formed but also because of those that are pruned. John Bruer (1997) described this developmental phenomenon as follows:

At birth, both nonhuman and human primate brains contain synapses that connect brain cells into circuits. Neonates have slightly fewer synaptic connections than do adults. However, early in postnatal development, the infant brain begins to form synapses far in excess of adult levels. This process of synaptic proliferation, called synaptogenesis, continues over a period of months that varies among species. This period of synaptic overproduction is followed by a period of synaptic elimination or pruning. This experience-dependent pruning process, which occurs over a period of years, reduces the overall number of synaptic connections to adult, mature levels, usually around the time of sexual maturity for the species. The mature nervous system has fewer synaptic connections than were present during the developmental peak. *It is the pattern, rather than simply the number, of these connections that form the mature brain's neural circuitry and that support normal brain function.* (p. 5; emphasis added)

Thus what is practiced becomes permanent, and only perfect practice makes perfect. Although this argument can and has been made purely from the psychology of learning processes, neural patterns get formed by learning and overlearning, not just from exposure.

Moreover, neural connections are never just cognitive. Environmental stimuli are processed, first, for whether they are life threatening. If so, they set in motion “fight or flight” reflexes. Then they are processed, connected, and conditioned to emotions, such as fear or happiness. Finally, though the total time may be only milliseconds, they are connected to long-term memories and cognitive processes to help interpret what the stimuli mean and what to do about them. Thus, for example, we not only master musical concepts by organizing our knowledge and performing the correct movement but we also associate feelings of love or patriotism, or what have you, to those songs: We play or listen with feelings that are part of those neural patterns. Or, for example, we associate math with threats, we become fearful, and we establish neural patterns supporting learned helplessness.³

The Empirical Base

In the third edition of the *Handbook of Research on Teaching*, Michael Dunkin (1986) reviewed the higher education literature and came to the following conclusion about Fred Keller's individualized (PSI) mastery system:

The single most significant conclusion to be reached from research on innovatory teaching methods in higher education is that the Keller Plan is clearly superior to other methods with which it has been compared. Indeed, the Keller Plan has been so consistently found superior that it must rank as the method with the greatest research support in the history of research on teaching. (p. 759)

Add Bloom's group-based (LFM) approach, and there have been hundreds of studies and dozens of reviews of those studies, including several meta-analyses.⁴ These reviewers consistently agree that the following are the active ingredients for the positive cognitive and affective outcomes of those studies:

1. Clear mastery objectives, properly sequenced for transfer of previous knowledge to current and future lessons
2. A preestablished, high passing standard
3. Grading that is criterion referenced, with corrections that encourage and require achievement of those high standards

Effect sizes in these meta-analyses run in the .4 to .6 range; that is, groups taught by mastery methods tend to do about half a standard deviation better than those taught by the traditional methods with which they were compared. We review much of this evidence in other ways in the Appendix, including explanations of meta-analysis and effect size. Of course, not all reviewers agree on the interpretation of those results, especially since effects are larger on criterion-referenced tests directly tied to the instructional domain than they are on standardized, norm-referenced tests. Thus Stallings and Stipek (1986) concluded that the gains are something like placebo or Hawthorne effects:

Mastery learning advocates no doubt make an important contribution to student learning by convincing teachers that all

children can master the curriculum. Indeed, the achievement gains for children in mastery-based programs may be explained as much by teachers' enhanced expectations, especially for the low-ability students, as by any other aspect of the program. (p. 746)

If this last conclusion were the correct one (to build on points 1-3 above), mastery learning has a fourth active ingredient: *Experience with mastery learning convinces teachers to believe that all children can learn the course objectives.* The standards have the goal of making teachers accountable for their students' academic achievement; therefore raising teachers' expectations so that they provide additional opportunity for their students to learn is no mean accomplishment.

There are other ways, fortunately, of discerning the meaning of mastery learning effects. Return, for example, to the effects on memory of attaining a high degree of original learning (see Section 2, "The Learning/Memory Base"). Among other things, memory should be greater for students who achieve mastery at original learning than for those who do not. This is also true for both fast and slow learners (see "Memory by Fast and Slow Learners" in the Appendix), which is perhaps not too surprising, because fast and slow learners are identified on tasks in the same domain (e.g., learning word lists or poetry). Suppose, however, that we ask how IQ relates to all of this. We already know, for example, that IQ is moderately but significantly correlated with memory. But suppose we randomly assign half the students to have to achieve a preset standard, while the other half (within the same IQ range) are exposed to the same material but do not have to achieve the preset standard. What happens to the correlation between IQ and surprise delayed-retention test scores?

A dissertation study on this very premise was completed recently, under the senior author's direction, by Marianne Baker (1999). The trait of intelligence was measured by the Cognitive Abilities Test (Thorndike & Hagen, 1986), yielding a composite IQ score as well as component scores for verbal, quantitative, and nonverbal IQ. Learning rate, fast versus slow, was measured by the Verbal Learning subtest of the Wide Range Assessment of Memory and Learning Test (WRAML) (Sheslow & Adams, 1990), and memory was measured by the Story Memory subtest on the WRAML. Specifically, for original learning, a short story was read aloud to fourth and fifth graders individually, immediately followed by a free-recall test on specific items

Table 1.2 Correlations Between Intelligence and Memory With and Without Mastery Standards

<i>Intellectual Trait</i>	<i>Memory After Exposure (N = 39)</i>			<i>Memory After Mastery (N = 42)</i>		
	<i>7 days</i>	<i>14 days</i>	<i>28 days</i>	<i>7 days</i>	<i>14 days</i>	<i>28 days</i>
1. Rate of learning	.36*	.43**	.31*	.11	.16	.16
2. Memory ability	.59**	.56**	.61**	.02	.05	-.02
3. Composite IQ	.49**	.59**	.57**	.06	.13	.19
4. Verbal IQ	.44*	.52**	.50**	-.02	.10	.11
5. Quantitative IQ	.42*	.48**	.45**	.11	.09	.15
6. Nonverbal IQ	.39*	.48**	.50**	.05	.16	.23

SOURCE: Adapted from Baker (1999), from Story C results (see endnote 4).

* $p < .05$, ** $p < .01$

of information as well as comprehension of ideas in the story. For the mastery group, this process was repeated until each student scored between 75% and 90% correct. The nonmastery group heard the story once and did the free-recall test. A week later, both groups were surprised with a written test of memory for the same items. Then students relearned under their respective conditions and finally were tested for retention again after 14 days and 28 days.

Table 1.2 shows the remarkable results regarding intellectual traits and memory.⁵ Under nonmastery conditions—that is, a single exposure for original learning, recall after 7 days, a single relearning opportunity, and then recall after 14 and 28 days—the correlations between intellectual traits and recall are all positive and significant. That is, higher-ability students tend to remember more, as society has come to expect.

In stark contrast, imposing a mastery standard of 75% to 90% correct on original learning and then again at relearning renders those standardized intellectual measures nonpredictors of how much is recalled: The correlations hover around zero and are all nonsignificant.

What mastery to a high standard can do, in summary, is virtually bypass the effects of IQ for specified educational objectives. What is recalled about educational lessons is more dependent on how well the material is mastered than on such traits as rate of learning or general intellectual abilities.

Discussion and Conclusions

Critics usually cite three points in opposition to mastery learning:

1. It helps slower students at the expense of the faster students.
2. It is too oriented toward basic knowledge and skills at the expense of creativity and higher levels of thinking.
3. It requires too much work of the teachers.

Let's consider each in light of the principles and theories previously enumerated. That fast learners are bored to tears waiting for slow ones to catch up is far too true of many educational programs, including badly implemented mastery programs. It is not true at all of Keller's (1968) individualized mastery plan, in which students complete course units at their own pace. For group-based mastery schemes, such as Bloom's (1971), it would be true only if mastery were misconceived solely as passing minimum competency tests, with no incentive for students to use their new competencies for higher-level intellectual purposes.

Thus, to counter Criticism #1, the following are needed:⁶

- *A grading system that earns a minimum passing grade—that is, a C or 70 for passing the initial mastery test with at least 75% to 80% correct.* Under the concept of mastery as a beginning rather than as an end state of learning, even a test score of 100% is just the initial phase of learning-forgetting curves. Thus it should earn an entry-level grade. Not passing the mastery test should have a grade of zero or "incomplete" attached to it so that, like driver's tests, initial mastery is conceived as an all-or-nothing affair: Either you get your license, or you do not.

- *A set of enrichment activities that use but go beyond the basic knowledge, skills, and principles required for mastery.* This includes reports on how these principles are applied in real life, creative projects and experiments, further readings or advanced problems to be solved, cooperative investigations or debates, and—most important—tutoring others (we really learn something well when we teach it). Such activities, because they provide overlearning, distributed practice, organization and construction of knowledge, and the like, earn bonus points when adequately completed: Add 5 to 10 points for each project to the minimum pass of 70, or move from C to B for one advanced project and A for two or three such projects.

Under such a grading scheme, the fast learners will no longer be rewarded with the highest course grades simply for beating out their slower peers. And because neither they nor their parents are likely to be satisfied with a minimal pass, they will have adequate incentive to go beyond initial mastery to earn a higher grade.

If enrichment activities as described above are included, then Criticism #2 is also refuted. To apply, analyze, and synthesize—and especially to tutor—concepts that were recently mastered is necessary and required in any sensible mastery scheme, and these are high-level skills indeed. Both fast and slow learners can be encouraged to extend themselves in such a system, whereas too many existing programs, both mastery and nonmastery, get bogged down doing only the basics.⁷

There is some truth in Criticism #3, that mastery learning is more work for teachers. Especially in the beginning, the teacher needs to (a) decide what is absolutely essential to be mastered, (b) create parallel forms of mastery tests, (c) invent activities and scoring keys/rubrics for mastery of performances and enrichment activities, (d) organize and order units or lessons to facilitate transfer of learning, and (e) publish and be able to defend the grading scheme, and so forth. Once those are complete—and such tasks can be shared with like-minded colleagues—they will also eventually be mastered and become routine. Then the hard work that remains encourages, indeed requires, each student to master the material and go beyond, which requires extra time providing feedback, conducting remediation and testing sessions, and encouraging students to maximize their potential. To us, that doesn't sound like *extra* work; rather, it is *teachers'* work.

Criticism #3 is often phrased in another way: "Teachers have too much material to cover; they simply don't have time for mastery learning." As demonstrated in previous sections, however, without mastery, more time is wasted because (a) there is little or no savings in relearning when the material is encountered again next month or next year, (b) the fast and slow learners grow farther apart because only the former will show savings, and (c) many students lose motivation, become learned helpless, and/or become alienated from school.

In conclusion, the current standards movement is long on rhetoric about every child learning to high standards, even so-called world class standards. Students and educators come to believe such mythology applies to them only when they are succeeding. Standards and how to teach to them are the bread and butter of mastery learning and criterion-referenced assessment, as documented above. The evidence and logic is there, derived both from successful and misconceived implementations. Will we learn from this accumulated evidence, or

will we, a generation from now, be speaking of the need for another “new math,” or new world-class standards? The following chapters explore these issues in more depth.

Summary

In this chapter, we have provided an overview of mastery learning and the principles on which it is based. Both a philosophy of instruction and a set of ideas for teaching and assessing, mastery learning requires that each student achieve at least minimal curriculum standards, then review and relearn them (in a spiral curriculum) at ever higher levels of thinking and extended applications. To do this requires that students be assessed in a criterion-referenced fashion, that is, without reference to other students but in direct relation to a specified set of instructional objectives and criteria for evaluating how well they have been attained. Teachers must also understand that what is originally learned is typically soon forgotten and therefore overlearning must be built into the curriculum. Fast and slow learners will benefit equally from such programs by spending the time they need to learn, by solving their identity crises, and thereby substituting earned competence for learned helplessness.

Notes

1. The learning/memory research is not explicit on exactly what the threshold for sufficient original learning is. Our estimate of 60% is probably not too far off, and it also has the virtue of being the lowest passing grade for most schools. Thus in practical memory terms, students who have achieved less than 60% correct at original learning will be mostly indistinguishable a few weeks later from students who have never been exposed to the material at all.

2. For example, Bugelski (1979); DeCecco (1968); Hulse, Egeth, & Deese (1980); Travers (1977). For a practical and readable example of the differences between massed and distributed practice in language learning, see Bloom & Shuell (1981).

3. We are fully aware that a controversy exists about whether there is a direct, logical link between a particular educational tactic and basic brain research (e.g., Bruer, 1999, vs. Brandt, 1999). We tend to side with Bruer (1997, 1999), believing that the inferences and applications usually drawn from brain-based research are the same as—and are more parsimoniously derived from—research on cognitive and behavioral processes. That is, the learning/memory curves and helplessness arguments made in the text were originally derived from specific psychological studies. Although recent

technology allows more complete understanding of the brain mechanisms that must necessarily undergird such generalizations, our understanding of the brain's activities vis-à-vis these psychological processes is still in its infancy. With such a caveat in mind, it is nevertheless interesting to speculate on teaching ideas that, in Brandt's (1999) words, may be "used to supplement what we know from other sources" (p. 238). Recent sources that may be useful in that spirit are Sousa (2001), Sprenger (2000), and Sylwester (2000).

4. For example, Anderson (1985); Anderson & Burns (1987); Arlin (1984); Block, Efthim, & Burns, (1989), Gentile (1997); Guskey & Pigott (1988); Kulik & Kulik (1991); Kulik, Kulik, & Bangert-Drowns (1990); Kulik, Kulik, & Cohen (1979); Martinez & Martinez (1992); Stallings & Stipek (1986).

5. Baker (1999) had a number of other conditions, including two different stories, "B" and "C." The results reported here are from Story C, which, unlike Story B (which had an unexpectedly restricted range of scores) provided a clear test of the hypothesis. Because the IQ measures intercorrelated as expected, and other correlations are in the ranges found by others (e.g., for the correlation between initial trials to relearn the correlation in Baker's study was .36, compared with .35 found by Gentile, Voelkl, Mt. Pleasant, & Monaco [1995] and .34 to .55 found by Stroud & Schoer, 1959), there is every reason to expect these correlations to be valid estimates for these analyses.

6. For other variations on these points for grading purposes, see Gentile (1997, pp. 481-489); Gentile & Murnyack (1989); Gentile & Stevens-Haslinger (1983); Gentile & Wainwright (1994).

7. See also Carroll's (1989) refutation of mastery learning as repetitive drill and practice of basics (p. 28), as well as Chapter 4 in this book.