General Article

NEW CONCEPTUALIZATIONS OF PRACTICE: Common Principles in Three Paradigms Suggest New Concepts for Training

By Richard A. Schmidt and Robert A. Bjork

We argue herein that typical training procedures are far from optimal. The goal of training in real-world settings is, or should be, to support two aspects of posttraining performance: (a) the level of performance in the long term and (b) the capability to transfer that training to related tasks and altered contexts. The implicit or explicit assumption of those persons responsible for training is that the procedures that enhance performance and speed improvement during training will necessarily achieve these two goals. However, a variety of experiments on motor and verbal learning indicate that this assumption is often incorrect. Manipulations that maximize performance during training can be detrimental in the long term; conversely, manipulations that degrade the speed of acquisition can support the long-term goals of training. The fact that there are parallel findings in the motor and verbal domains suggests that principles of considerable generality can be deduced to upgrade training procedures.

Over the past several years, through the normal process of conducting our own individual research programs (in movement learning and human memory, respectively), and as a consequence of listening to and reading reports of each other's work, we have repeatedly encountered research findings that seem to violate some basic assumptions about how to optimize learning in real-world settings. For example, increasing the frequency of information presented to learners about performance errors during practice improves performance during training, yet can degrade performance on a test of long-term retention or transfer. Increasing the amount of task variability required during practice, in contrast, depresses performance during training, yet facilitates performance on later tests of the ability to generalize training to altered conditions. Such findings challenge common views of skill learning. Compared with some baseline training condition, how can a factor that enhances performance in practice interfere with retention or transfer performance? Even more intriguing, how can another factor that degrades performance in practice enhance retention performance?

These findings—and others we discuss below—are obtained from diverse research paradigms that employ several different verbal and motor tasks, and the theoretical motivations guiding those research efforts are often different as well. Taken together, however, these findings suggest that certain conceptualizations about how and when to practice are at best incomplete, and at worst incorrect. These findings also have some theoretical implications with respect to the processes involved in practice, particularly as they relate to the acquisition of realworld skills.

In this article, we first describe what we regard as some of the viewpoints, assumptions, and paradigms that, implicitly or explicitly, have provided the foundation for the typical procedures that guide practice and skill acquisition. These views of learning, though flawed in our opinion, have had a strong influence on the design of learning environments in educational, industrial, and military contexts. We then illustrate those flaws with examples from three different research paradigms, and we argue for a set of processes occurring during practice that can, at least in general terms, account for such findings.

SOME COMMON ASSUMPTIONS ABOUT PRACTICE

When researchers conduct studies of practice and learning, they generally ask learners to engage in practice at some task in an *acquisition phase*, and some independent variable is manipulated. The independent variable of interest can be of various types, such as the nature of instructions, the type of feedback, or the scheduling of practice, and the performance on some task is typically charted as a function of practice trials for groups operating with different levels of this variable. The logic of such paradigms, of course, is that those acquisition conditions

that speed the rate of improvement, or cause subjects to reach criterion more quickly, or in general result in more effective performance in practice, are expected to be the most effective for learning this particular task. Learning, after all, can be indexed by the improvements in skill across practice; it seems unavoidable, therefore, to conclude that those conditions in acquisition that speed gains in performance have done so because they have enhanced the processes of learning in some way. There are two related problems with this view of the learning process.

Problem 1: Acquisition Performance Is an Imperfect Indicator of Learning

In the recent era of research on the processes of learning, memory, and performance, researchers seem to have lost track of a critical distinction between the momentary strength or accessibility of a response and the underlying habit strength of that response. The major learning theorists of an earlier era recognized decades ago that experimental variables applied during training can have two distinct kinds of effects (see, e.g., Estes, 1955; Guthrie, 1952; Hull, 1942; Skinner, 1938; Tolman, 1932). First, of course, such variables can have the relatively permanent effects that are the usual focus when learning is examined. That is, these variables might speed the development of some relatively permanent capability for responding (the usual definition of learning, and the one we use here), so that a group of subjects with more of this capability will usually perform more effectively during practice than a group with less of this capability. Second, however, there may also be temporary effects of such experimental manipulations-effects that exaggerate or diminish performance differences while the variables are operating, with these performance differences vanishing or being markedly altered as soon as the subjects are allowed to rest, or when the manipulation is removed. Such performance effects can be mediated by a host of factors, such as the elevating effects of motivational instructions or the administration of feedback, as well as the depressing effects of physical (or mental) fatigue and boredom. A given experimental manipulation can have either or both of these learning and performance effects.

This important distinction has been mostly ignored since the late 1950s (see, e.g., Salmoni, Schmidt, & Walter, 1984, in the area of feedback and skill learning), and it is interesting to speculate why that might be the case. In our opinion, the information processing metaphor, which has dominated much of the modern era of research, has led theorists away from such a distinction. That metaphor, based as it is on the architecture of the typical digital computer, does not readily suggest the kind of dual memory representation implied by, for example, habit strength and reaction potential (Hull, 1943). (For more on these and related arguments, see Bjork, 1989, and Bjork & Bjork, 1992.)

For present purposes, the important point is that only certain kinds of performance changes can qualify for the label *learning effects*. For us to agree that one level of some variable has produced more learning than another, we usually demand that these differences have some permanence across time, or that the differences be able to survive the removal of the manipulation in question. The problem is to discover which of many possible practice variables produce learning effects in the sense just defined—that is, to determine whether a given independent variable has effects that are relatively permanent or are merely transitory.

Testing posttraining retention and transfer

The standard approach to this problem is to use various kinds of transfer or retention tests as a means of evaluating the extent to which true learning has taken place. Assume that two groups of subjects practice under different levels of some independent variable during an acquisition phase. For example, they might be learning foreign vocabulary words, with frequency of feedback being manipulated (after every trial vs. after every fifth trial). Differences between groups during the acquisition phase could reflect differences in learning or performance (or both). It is critical, therefore, to add a retention phase (sometimes called a *transfer phase*), conducted after an interpolated interval that is long enough to ensure that any temporary effects of the independent variable have been dissipated. If subjects are then tested on the same (or similar) task again under equated levels of the independent variable (so that differential temporary effects cannot reappear across trials), relative performance differences between the two groups can be viewed with some confidence as reflecting differential learning that occurred during the acquisition phase.

Special considerations in real-world training

Measuring the actual level of learning that results from a training regimen of some kind may not seem to be a particularly serious problem for scientists or scientifically trained professionals involved in training, as these issues have been (or should have been) familiar to us for several decades. But the problem is far more serious for the typical person who is actually doing the training in some real-world setting. Here, it is easy to imagine that trainers would make every effort to adjust the training (measured as either speed of acquisition, that is, the trials or time necessary to reach some specified performance goal, or the level of performance achieved after a fixed amount of training time or trials). Without even giving the

matter much thought, trainers might easily assume that maximizing performance during training is their major goal; trainers may themselves even be evaluated in terms of their trainees' performance during training.

Two other considerations exacerbate the problem in real-world settings. First, while instructors have ample opportunity to view their students during practice, they frequently do not have a chance to examine their learners on the transfer or retention tests that are the real goal of training. Such posttraining performance is often delayed or in a different location than the original training. Second, instructors can also be misled by their own trainees: In a study of learning keyboard skills under different practice schedules (Baddeley & Longman, 1978), for example, the schedules that were most preferred by subjects produced the least learning.

Problem 2: Acquisition and Retention Phenomena Are Not Separable

Learning processes versus retention processes

Our basic argument is that relative amount learned should be measured by performance on retention tests of various kinds, and that performance levels in acquisition are "flawed," or at least ambiguous, with respect to the amount learned. Note that this is quite a different view from that often taken in educational and training settings, where learning and retention are seen as two different phenomena. "Learning" is assumed to refer to that set of processes occurring during the actual practice on the tasks of interest, as assessed by performance measures taken at that time, whereas "retention" is seen to involve the set of processes that occur after practice is completed, during some retention interval, and prior to a retention test. Because learning and retention are thought to be different phenomena, they tend to be studied with separate methods, by different scientists, and even in different laboratories. Rather than viewing learning and posttraining retention as separable phenomena, however, we argue that the effectiveness of learning is revealed by, or measured by, the level of retention shown.

Criteria against which training should be evaluated

In most educational, military, and industrial settings, the effectiveness of a training program can be evaluated by several criteria, depending on what we would like our learners to be able to do. Certainly, one of the most important of these is posttraining performance; we want trainees to be able, many months after the training program is completed, to perform well, or at least adequately. This criterion is especially important in times of natural disasters and man-made emergencies, when key people must perform critical functions in situations that reoccur, typically, only after very great delays. A crisis in a nuclear power plant would be a prime example. This criterion is also important in minimizing the time and money spent on retraining or refresher courses.

Another criterion is generalization. Whereas it is important to be able to perform the specific skill acquired in practice some months later in a retention test, it is also important to be able to generalize to variations of that skill, perhaps to be performed in contexts different from those experienced in acquisition. For example, the trainee might have to generalize the skill acquired under quiet, controlled conditions in a classroom to a noisy, hot, and cluttered environment in the workplace. The capability to perform in the presence of stress, sleep loss, or fatigue may be critical in some situations, and the need to perform a simultaneous secondary task effectively may be important as well. There may also be a need to have learning generalize to other learning environments, allowing new tasks to be learned more quickly and easily. The acquisition condition that is most effective-given these criteria-is the one leading to the highest performance on a novel version of the task, or on a task performed or practiced under novel conditions. Thus, rather than thinking of learning and generalizability as separate concepts—as is often done—we interpret the capability to generalize as one measure of learning, and as a basis for selecting among various training conditions.

It is perhaps not new to suggest that there are several goals of training and instruction—such as long-term retention, generalizability, and resistance to altered contexts. What is new, however, is the notion that the training conditions to achieve these training goals are not necessarily those that maximize performance in the acquisition phase. In fact, as we show next, there can be conditions for which the effectiveness of training—as measured by one or another of these alternative criteria is best achieved by a condition that produces relatively poor performances during training.

INTRODUCING DIFFICULTIES FOR THE LEARNER CAN ENHANCE TRAINING: THREE ILLUSTRATIONS

In this section, we discuss three broad situations in which, relative to a "standard" practice condition, some condition in acquisition that slows the rate of improvement or decreases performance at the end of practice nonetheless yields enhanced posttraining performance. One of these examples involves variations in the way tasks can be ordered for practice, with the focus on the criterion of producing effective skill retention. A second example involves variations in the nature and scheduling of feedback for learning, again with the emphasis on enhancing a retention criterion. Finally, a third example involves inducing variation among versions of the tasks

to be practiced, with the focus on a criterion of generalizability.

Scheduling of Tasks During Practice

Consider the general problem in which several different tasks or items are to be learned in a practice session of a fixed length. How should the practice on these tasks or items be organized to maximize learning and retention?

Experiments with motor tasks

Many variations in practice scheduling are possible. Shea and Morgan (1979) contrasted random and blocked schedules of practice, two schedules that differ substantially in terms of what Battig (1966) referred to as "contextual interference." In Shea and Morgan's study, blocked practice involved sequential trials at Task 1, Task 2, and Task 3, with all trials for a given task being completed before moving on to the next. Random practice, in contrast, involved the same number of trials at the three tasks, but the order was randomized so that a given task was never practiced on successive trials. Thus, blocked practice resembles what we usually term drill. The tasks required rapid, multiple-component arm movements, with the goal of minimizing response time, and different tasks had different patterns. After practice in an acquisition phase, retention tests were given after 10 min and 10 days. These retention tests were given under either random or blocked conditions. The experiment, therefore, was designed to assess the effect of random versus blocked practice on performance measured under blocked or random conditions.

The results are shown in Figure 1. During the acquisition phase, at the left, there was a clear advantage for the subjects who practiced under the blocked conditions, especially in the initial phases of practice, but continuing until the last acquisition block. Amount of learning, however, as measured by the tests of posttraining retention, tells a different story. Consider first the tests given under the random conditions, shown as the filled and open squares. There was a strong advantage for retention for the subjects who practiced under the random conditions in acquisition. That is, even though the random conditions were less effective during the acquisition phase, they were better than the blocked conditions on the random retention test. These differences are especially impressive given the ecological validity of random tests; that is, most real-world behaviors are not produced in blocked contexts.

An alternative interpretation of the advantage of random practice for random retention conditions is that the practice conditions in the acquisition and test phases were identical for the random subjects, but were different



Fig. 1. Performance on movement speed tasks under random (R) and blocked (B) conditions in acquisition and, after 10 days, in retention tests under random or blocked conditions; in retention, the first letter indicates the acquisition condition, and the second represents the retention condition. Redrawn from Shea and Morgan (1979).

for the blocked subjects. This is, in effect, a kind of "identical elements," specificity-of-learning, or similarity argument. This relatively uninteresting interpretation cannot, however, explain the retention performance observed under the blocked conditions (shown as open and filled triangles). Once again, there was an advantage though much reduced—for the subjects who practiced under random conditions during acquisition.

Regardless of whether the retention test was itself random or blocked, then, it was always more effective to have practiced under random conditions than under blocked conditions. Remarkably, this was the case even though the random condition was detrimental to performance during acquisition. Relative to blocked practice—a schedule that most people would feel was "natural" or optimal—random practice is, then, a first example of a manipulation that degrades performance in acquisition, yet enhances performance at retention and contributes to the capability to perform in different contexts (see also Wulf & Schmidt, 1988).

Similar effects have been found in several other experiments using real-world skills (serving in badminton, Goode & Magill, 1986; keyboard skills, Baddeley & Longman, 1978), as well as laboratory tasks (Lee, 1988; Lee & Magill, 1983). One exception is that at minimal levels of practice, blocked practice produces better retention than does random practice, but this effect is reversed with additional levels of practice (Shea, Kohl, & Indermill, 1990). These phenomena, and various theoretical interpretations thereof, have been reviewed recently by Magill and Hall (1990).

Experiments with verbal tasks

The effects of blocked versus random practice in the motor skills literature are analogous to certain verballearning phenomena typically studied under the heading of spacing effects (Melton, 1967). Here, the general problem is that distinct items presented serially are to be learned, and the question is how the study trials on a given item should be interleaved with the study trials on other items to generate maximal retention. In general, spacing of repetitions yields better long-term retention than does massing of repetitions-often much better. If the final retention interval is short, however, massed repetitions can yield better performance than spaced repetitions. (For examples of such interactions involving intervals ranging from seconds to minutes to days. respectively, see Peterson, Hillner, & Saltzman, 1962; Glenberg, 1979; Glenberg & Lehmann, 1980; and Bahrick, 1979.) The interaction of spacing interval and retention interval may again mislead people responsible for training; on the basis of performance during acquisition alone, massed repetitions may appear to be superior to spaced repetitions.

In a variety of real-world situations, the question is not how one should distribute the repetitions of items, but rather how one should distribute one's effort to practice the retrieval of those items. Two experiments (Landauer & Bjork, 1978) examined how such retrieval efforts should be scheduled to optimize long-term retention. In the first experiment, subjects were asked to learn a number of names of hypothetical people. During the study phase, a given name was presented once and then tested three times (by presenting the first name as a cue for the last name or the last name as a cue for the first name). The intervals from the initial presentation of a given name to each successive test of that name were filled with different numbers of intervening presentations and tests of other names. Following the study phase, there was a 30min retention interval filled with a distracting activity prior to a final retention test for all the names.

Two aspects of the results of this experiment are of interest. First, as shown in Figure 2, the conditions that yielded optimal performance on the tests during acquisition yielded the poorest long-term retention. In a condition with 0 items intervening between successive tests, performance on those tests averaged about 95%, but performance dropped to 33% on the final retention test. Other uniform-spacing conditions, with 4 or 5 intervening items between successive tests, yielded poorer performance during acquisition (about 43% correct), but better final retention (41% correct). Second, an expanding sequence of intervals prior to each successive test on a given name during acquisition (0, 3, and 9 intervening items, or 1, 4, and 10 items) appeared to yield optimal retention performance (48% correct).



Fig. 2. Percentage correct name recall on the third of three tests embedded in a study phase and on a 30-min delayed retention test, as a function of the spacing of the three study-phase tests. Redrawn from Landauer and Bjork (1978).

In the second experiment, subjects were asked to memorize a first and last name corresponding to each of a set of facial photographs. During the acquisition phase, after an initial pairing of a given name and face, there were four subsequent tests of that face-name combination, each of which consisted of presenting the face and the first (or second) name as a cue for the missing name. The intervals separating the successive tests of a given name formed an expanding sequence (0, 1, 3, and 8 intervening events) or a uniform sequence with the same average interval length (3, 3, 3, and 3 intervening items), and there was again a test of final retention after a 30-min delay. On the final test, each face was shown alone, and subjects were asked to recall both the first and last name corresponding to that face.

Once again, an expanding sequence was more effective than a uniform sequence for long-term retention. In fact, for the expanding condition, the retention of a name presented only once (and tested four times) was greater than retention of a name presented-together with a given face-five times (66% vs. 58% correct). This result illustrates the general principle that tests are potent learning events-often more potent than presentationsparticularly when the tests are difficult enough to constitute a type of retrieval practice with respect to the criterial retention test. There are analogous results in the motor memory literature. For example, Hagman (1983), using an arm-positioning task, found that four test trials that involved attempting to repeat a once-presented position were more effective for retention than were four presentation trials in which the subject moved to a stop defining the target position.

In terms of the goal of enhancing long-term retention,

expanding-interval retrieval practice may well be an important component of an optimal training program. Rea and Modigliani (1985), for example, have gone on to show that expanding retrieval practice is about twice as effective as massed practice in children's memorization of multiplication facts and spelling words. Such effects of expanding-interval retrieval practice in the verbal domain seem quite closely related to another effect we discuss in the next major section—namely, the scheduling of the number of practice trials between presentations of feedback in skill learning.

Common principles

In each of the foregoing paradigms, the condition that produced the best retention performance seemed to have the characteristic that it provided added "difficulty" for the learner during the acquisition phase, reflected in poorer performance at that time. Thus, as we view it, random practice serves to keep the performer from generating a stable "set" for a particular task, and forces the learner to retrieve and organize a different outcome on every trial. Similarly, the spacing of repetitions may prevent superficial massed rehearsal.

These notions suggest that retrieval practice (Bjork, 1975, 1988), in which the learner is actually given practice at the process of retrieving information from memory, may be an important factor in all of these paradigms. Indeed, other information processing activities that cause forgetting of the to-be-remembered information, and thus require practice at retrieving it again on a subsequent trial, are beneficial for retention (Bjork & Allen, 1970; Cuddy & Jacoby, 1982).

We view retrieval practice as a specific case of transfer-appropriate processing (Morris, Bransford, & Franks, 1977); practice at retrieving in acquisition is appropriate for the need to retrieve during retention or transfer tests. Consistent with this view, there are results in the literature (Allen, Mahler, & Estes, 1969; Hogan & Kintsch, 1971) suggesting that tests as learning events relative to presentations as learning events—become more effective as (a) the retention interval preceding a criterion test is increased and (b) the criterion test stresses recall rather than recognition.

It is clearly too extreme to argue that every manipulation causing difficulty for the learner during practice will enhance retention performance (see, e.g., Shea & Upton, 1976, who showed that interpolated processing tasks degrade performance both during practice and on retention tests); but if the manipulation demands other kinds of information processing—such as retrieval practice—that are also needed for retention performance, then such added difficulty can be expected to enhance retention performance.

Feedback During Skill Acquisition

A second illustration involves the nature and scheduling of feedback presented to learners during an acquisition phase. It has generally been understood that any variation of feedback in practice that makes the information more immediate, more accurate, more frequent, or more useful for modifying behavior will contribute to learning, as measured during the acquisition phase. This view of the relationship of feedback and learning has served as the basis for instructional practice in many settings, as well as for the design of simulators. Recent evidence, however, suggests that this generalization must be qualified.

Experiments with motor tasks

In one study (Schmidt, Young, Swinnen, & Shapiro, 1989), subjects were asked to learn a relatively complex arm movement in which the subject was to produce two reversals in direction such that the time of the action was as close to a set goal as possible. In one condition, feedback about the movement-time error was given after each trial, a more or less standard schedule typically thought to optimize learning. Feedback was also given in summary form (see Lavery, 1962), in which the subject received feedback about each of a set of trials (e.g., 5) only after the last trial in the set was completed. This feedback was given in the form of a graph of performance against each of the trials in the set, so that the subject could see the error on each of the previous trials. The summary length—the number of trials summarized on the graph was either 1 (the every-trial feedback condition mentioned earlier), 5, or 15 trials. After practice under these conditions in an acquisition phase, subjects were given tests of posttraining retention (without any feedback) after 10 min and 2 days.

The results of this experiment are shown in Figure 3. In the acquisition phase, subjects in the 1-trial summary condition performed more accurately throughout practice than the other groups, with generally larger errors being produced as the summary length increased. It is clear that increased summary length interfered with performance during training, both in slowing the rate of approach to the asymptote and in generating larger errors near the end of practice. However, when performance was evaluated on the delayed retention test, the most effective performance was generated by the 15-trial group, with generally increasing errors as the summary length in the acquisition phase decreased. That is, there was a clear negative relationship between the level of performance in acquisition and the level of performance in retention. These data tend to contradict the long-held view that making feedback more useful is effective for learning, as the 15-trial condition seemed to provide dif-



Fig. 3. Mean errors in a movement-patterning task for three different summary-feedback lengths in acquisition, and on no-feedback retention tests given after 10 min or 2 days. Redrawn from Schmidt, Young, Swinnen, and Shapiro (1989).

ficulties in relating the feedback received in the graph to the error on the trial to which it referred (see also Schmidt, Lange, & Young, 1990).

Similar effects were obtained when feedback was given in acquisition either on every trial (100% condition) or on only half of the trials (Winstein & Schmidt, 1990; see also Wulf & Schmidt, 1989). In the latter, 50%, condition, the feedback was faded, such that feedback was given on every trial early in practice and gradually withdrawn across practice. Retention performance was measured after 10 min and 2 days, either with or without feedback being presented (in separate experiments). In both experiments, the 100% and 50%-faded groups were essentially similar in the acquisition phase. But on the posttraining retention tests, the 50%-faded group had more effective performance, with the differences becoming larger as the retention interval increased. These data contradict traditional views of feedback operation in that providing half the number of feedback presentations in acquisition produced more effective retention performance. The general finding that expanded spacing of feedback presentations enhanced retention is analogous to the finding (Landauer & Bjork, 1978) that expanded spacing of repetitions was more effective for name learning, suggesting again that some common features underlie these two paradigms.

Experiments with verbal tasks

During the 1960s, a dozen or so paired-associate experiments were conducted in which the proportion of responses that received feedback was manipulated during the acquisition phase. The percentage of occurrence of response members (% ORM) was defined by the percent-

age of trials on which the correct response term was shown after the subject had responded to the stimulus term. All of these studies, unfortunately, have characteristics that prevent them from being compared directly with the work just mentioned on motor behavior. For example, practice was always provided until a particular criterion was reached (e.g., 100% correct); because improvement in acquisition was faster with more frequent feedback, this procedure confounded the percentage feedback in acquisition with the amount of practice. Also, delayed retention tests were never given, which is unfortunate in view of the motor findings that the benefits of infrequent feedback seem to increase with the longer retention intervals (see, e.g., Fig. 3). Even so, several of these studies suggest that reducing the percentage feedback in acquisition-in some cases from 100% to 0%has negligible effects on performance in immediate retention (Krumboltz & Weisman, 1962; Schulz & Runquist, 1960), suggesting a rough parallel to the work in motor behavior.

Recently, Schooler and Anderson (1990) examined feedback frequency effects in learning the computer language LISP, showing that (relative to frequent feedback) decreasing the number of feedback presentations depressed performance in acquisition, but facilitated retention performance. This work suggests that these effects might be generalizable to a variety of cognitive activities as well as to the motor behaviors discussed in the previous section.

Common principles

One interpretation of this work is that frequent feedback during the acquisition phase provides several advantages, one of which is the guidance toward the correct behavior. But it also provides some disadvantages (see Schmidt, 1991a). One possibility is that frequent feedback comes to be a part of the task, so that performance is disrupted in retention when the feedback is removed or altered. Also, frequent feedback could block information processing activities that are important during the acquisition phase for acquiring the capability to produce effective performance at retention. One possibility is that frequent feedback blocks the processing of responseproduced (kinesthetic) feedback, leading to less effective error-detection capabilities for use in retention (Schmidt et al., 1989). Another possibility is that frequent feedback makes performance too variable during practice, preventing the learning of a stabilized representation of the kind necessary to sustain performance on a later retention test.

Notice that, except for the particular terminology used, these accounts are very similar to those offered with respect to the spacing paradigms in the previous section. The general point is that certain "difficult" train-

ing conditions may foster various kinds of processing activities that are required for effective retention performance.

Induced Variability of Practice

A final example involves the intentional variation, along a single dimension, of the task to be learned in acquisition. In this case, the criterion test performance typically requires performance on some novel variation not experienced in the acquisition session. The question is whether this intentional variation during practice, versus a consistent practice schedule, is effective for transfer to some novel retention test.

Experiments with motor tasks

Numerous experimenters have dealt with this issue, but Catalano and Kleiner (1984) made the point very well. They used a coincident-timing task in which subjects responded to a simulated moving object by pressing a button when it reached a predefined coincidence-point. Subjects received either constant practice at one target speed (either 5, 7, 9, or 11 mph) or variable practice at all four of these speeds for the same number of total trials. Learning was evaluated on a retention test in which novel speeds that lay outside the range of the subjects' previous experiences were presented (1, 3, 13, and 15 mph).

In acquisition, performance in the variable condition was generally less accurate than performance in the constant condition (52 vs. 38 ms absolute error, on average), perhaps reflecting the common view that performing one thing repeatedly is generally more effective than performing four different things. But results for the retention test of generalization to novel speeds, shown in Figure 4, show the variable group was more accurate than the constant group.

Many other experiments in the motor skills literature demonstrate similar findings (see Shapiro & Schmidt, 1982, for a review), with especially strong effects for children. For example, Kerr and Booth (1978) had 8-year-old subjects toss beanbags to targets 2 ft and 4 ft away (variable group) or only to a target 3 ft away (constant group). On a subsequent test using the 3-ft target—the distance practiced by the constant group, but never practiced by the variable group—the variable group performed with greater accuracy than the constant group. This result suggests that learning how to modulate the relationships among the target distances was more important for a test at any one target than was specific experience, even at the particular target distance used at test.

This collection of results about variable practice is usually interpreted in terms of schema theory (Schmidt, 1975; Wulf & Schmidt, 1988). The idea is that practice variability enhances the effectiveness of rules (schemata)



Fig. 4. Mean timing error for transfer tests to novel target speeds after variable or constant practice in the acquisition phase. Redrawn from Catalano and Kleiner (1984).

that relate the external task requirements to the internal movement commands. But in terms of the arguments raised in the present article, these experiments suggest that variable practice alters the practice context to force a change in behavior from trial to trial, encouraging additional information processing activities about the lawful relationships among the task variants. The result is learning that contributes to performance on the test of retention or generalizability, even though these activities detract from momentary performance during the acquisition phase.

Experiments with verbal tasks

Several investigations in the concept formation literature provide analogous findings to those seen in the motor literature. For example, Nitsch (1977; see Bransford, Franks, Morris, & Stein, 1979) had subjects learn novel concept-words (e.g., to "crinch" was to offend someone) by providing several uses of the word that were in either a constant context (all in a restaurant) or a variable context (in numerous settings). Constant contexts were more effective than variable contexts for enabling subjects to identify the concept in the same context as it was presented earlier, and were probably more effective in the acquisition phase as well. However, when the subjects were asked to recognize novel examples of the concept, variable practice was more effective than constant practice.

With a different paradigm, Mannes and Kintsch (1987) asked subjects to study a passage of text, preceded by an outline that was in either the same or a different organization as the text materials. The different-organization outline can be thought of as a kind of variable practice, and the same-organization outline a form of constant

practice. When the subjects were asked to recall the original text materials, the same-organization outline was more effective. But when the subjects were asked to do creative problem-solving tasks that required a deeper understanding of the text materials, the different-organization outline was more effective.

Both of these examples, together with the motor examples discussed above, suggest that even though constant practice may lead to more effective performance in the acquisition phase, and often more accurate verbatim recall of the materials presented, constant practice produces less effective capabilities to generalize knowledge to novel situations than does variable practice.

CONCLUDING COMMENTS

A fundamental concern here has been the characterization of learning, its measurement, and the interpretations that are to be drawn from investigations of acquisition phenomena. Learning is obscured during the acquisition (or practice) phase because relatively permanent effects may be confounded with temporary performance effects that disappear quickly after the practice session is finished, or when the test conditions are changed. We advocate, therefore, the use of various kinds of transfer or retention tests on which (and only on which) the relatively permanent effects of the conditions in acquisition are evaluated. We have provided three experimental variations of practice in which conditions that facilitate performance during the acquisition phase are ineffective for learning as measured on a retention or transfer test. In each of those cases, there appear to be analogous effects across markedly different motor- and verbal-learning paradigms.

We are struck by the common features that underlie these counterintuitive phenomena in such a wide range of skill-learning situations. At the most superficial level, it appears that systematically altering practice so as to encourage additional, or at least different, information processing activities can degrade performance during practice, but can at the same time have the effect of generating greater performance capabilities in retention or transfer tests. If these processing activities are selected so that they are also needed for success at a test of retention or generalizability, then such conditions will facilitate learning.

What are the processes underlying these empirical effects? We have only begun to ask this question, and answers are necessarily very tentative at present. Many possible information processing activities have been postulated in the different tasks and paradigms mentioned here, such as the need to retrieve information that has faded from memory in name learning, the need to evaluate one's own response-produced feedback in motor learning, and the need to associate various different facts or actions into a single concept or schema. Other such processes have been suggested as well, and each of these paradigms has an active literature in which these various possibilities are argued and contrasted.

This perspective is distinct from the earlier viewpoints about the specificity of encoding (Tulving & Thomson, 1971) or specificity of abilities (Henry, 1958/1968), in which the overlap of the objective acquisition and test conditions is the critical variable for learning. Whereas this overlap is undeniably of some importance for test performance, there is ample evidence presented here and elsewhere that this is not the only factor, and perhaps is not even the major factor, for test performance: For example, if the test is given under a blocked condition, random practice in acquisition is more effective for this test than is practice with the identical blocked condition (Shea & Morgan, 1979; see Fig. 1 here). Also, even if the test is given under 100% feedback conditions, a 50%faded condition in acquisition is better at test than practice under the identical 100% condition (Winstein & Schmidt, 1990, Experiment 3). Finally, when test performance requires a beanbag toss of 3 ft, varied practice at 2- and 4-ft distances is better than practice at the identical 3-ft distance. All of these examples-taken from each of the paradigms mentioned here-tend to violate the specificity view that the simple overlap of conditions between acquisition and test contexts determines test effectiveness.

We prefer to suggest that the more important principle is the overlap of the processes necessary for performance at the test and the processes practiced during acquisition, refined from the ideas of transfer-appropriate processing (Bransford et al., 1979). Note that the overlap of relevant processes does not necessarily mean that there is overlap of the objective conditions of practice, as we have shown here several times already. If certain acquisition conditions force the learner to engage in processes that are also critical for test performance, then those conditions will be judged as effective for learning (because they facilitate test performance), even though they may exhibit different superficial conditions. Also, these conditions that maximize learning may not be very effective for performance during the acquisition phase, as they provide various "difficulties" for the learners. Random practice, reduced feedback, and variable practice all degrade performance during practice relative to more "ideal" conditions in acquisition, yet all can be argued to exercise information processing activities that are critical for performance at the test. In other words, these conditions can be considered as effective for learning because they prepare the learner for the processing that will be required at test.

Certainly, then, no single type of extra information processing activity will be expected to underlie all of the

tasks and paradigms discussed here. Even so, these data suggest a new conceptualization, or framework, for learning and training that has broad implications for educational practice (see, e.g., Christina & Bjork, 1991). From a practical perspective, this framework would stress that a trainer's major goal is to focus clearly on the criterion performance, and to understand what kinds of processes are required for its proficiency. Then, practice activities that exercise these particular processes could be designed (see, e.g., Schmidt, 1991b, chap. 11). The criterial version of many tasks, for example, involves the execution of an essentially novel response that cannot have been practiced previously, such as the solution of a particular mathematical word problem on the job, or the execution of a basketball shot from a location never before experienced. In such cases, practice could be organized in a way to facilitate transfer and generalization, and a form of variable practice would be recommended. Other practice conditions would optimize performance in other contexts, as we have argued here.

Acknowledgments—This article was based in part on a concept paper submitted to the U.S. Army Research Institute (ARI), Basic Research (Schmidt & Bjork, 1989), and was supported in part by Contract No. MDA 903-85-K-0225 from ARI to R.A. Schmidt and D.C. Shapiro. Thanks to William Estes and Tim Lee for valuable comments on an earlier draft, and to Jack Adams for pointing us to the research on % ORMs.

REFERENCES

- Allen, G.A., Mahler, W.A., & Estes, W.K. (1969). Effects of recall tests on long-term retention of paired associates. Journal of Verbal Learning and Verbal Behavior, 17, 573-585.
- Baddeley, A.D., & Longman, D.J.A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, 21, 627-635.
- Bahrick, H.P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. Journal of Experimental Psychology: General, 108, 296–308.
- Battig, W.F. (1966). Facilitation and interference: In E.A. Bilodeau (Ed.), Acquisition of skill (pp. 215-244). New York: Academic Press.
- Bjork, R.A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), Information processing and cognition: The Loyola Symposium (pp. 123-144). Hillsdale, NJ: Erlbaum.
- Bjork, R.A. (1988). Retrieval practice and the maintenance of knowledge. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), *Practical aspects of memory II* (pp. 396-401). London: Wiley.
- Bjork, R.A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H.L. Roediger & F.I.M. Craik (Eds.), Varieties of memory and consciousness: Essays in honor of Endel Tulving (pp. 195-210). Hillsdale, NJ: Erlbaum.
- Bjork, R.A., & Allen, T.W. (1970). The spacing effect: Consolidation or differential encoding? Journal of Verbal Learning and Verbal Behavior, 9, 567– 572.
- Bjork, R.A., & Bjork, E.L. (1992). A new theory of disuse and an old theory of stimulation fluctuation. In A.F. Healy, S.M. Kosslyn, & R.M. Shiffrin (Eds.), From learning processes to cognitive processes: Essays in honor of William K. Estes, (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum.
- Bransford, J.D., Franks, J.J., Morris, C.D., & Stein, B.S. (1979). Some general constraints on learning and memory research. In L.S. Cermack & F.I.M. Craik (Eds.), Levels of processing in human memory (pp. 331-354). Hillsdale, NJ: Erlbaum.

Catalano, J.F., & Kleiner, B.M. (1984). Distant transfer in coincident timing as a function of variability of practice. *Perceptual and Motor Skills*, 58, 851-856. Christina, R.W., & Bjork, R.A. (1991). Optimizing long-term retention and transfer. In D. Druchman & R.A. Bjork (Eds.), In the mind's eye: Enhancing human performance (pp. 23-56). Washington, DC: National Academy Press.

- Cuddy, L.J., & Jacoby, L.L. (1982). When forgetting helps memory: Analysis of repetition effects. Journal of Verbal Learning and Verbal Behavior, 21, 451-467.
- Estes, W.K. (1955). Statistical theory of distributional phenomena in learning. Psychological Review, 62, 369-377.
- Glenberg, A.M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7, 95-112.
- Glenberg, A.M., & Lehmann, T.S. (1980). Spacing repetitions over 1 week. Memory & Cognition, 8, 528-538.
- Goode, S., & Magill, R.A. (1986). The contextual interference effects in learning three badminton serves. Research Quarterly for Exercise and Sport, 57, 308-314.

Guthrie, E.R. (1952). The psychology of learning. New York: Harper and Row.

- Hagman, J.D. (1983). Presentation- and test-trial effects on acquisition and retention of distance and location. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 334-345.
- Henry, F.M. (1968). Specificity vs. generality in learning motor skill. In R.C. Brown & G.S. Kenyon (Eds.), Classical studies on physical activity (pp. 331-340). Englewood Cliffs, NJ: Prentice-Hall. (Original work published 1958)
- Hogan, R.M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 10, 562-567.
- Hull, C.L. (1943). Principles of behavior. New York: Appleton-Century-Crofts.
- Kerr, R., & Booth, B. (1978). Specific and varied practice of a motor skill. Perceptual and Motor Skills, 46, 395-401.
- Krumboltz, J.D., & Weisman, R.G. (1962). The effect of intermittent confirmation in programmed instruction. Journal of Educational Psychology, 53, 250-253.
- Landauer, T.K., & Bjork, R.A. (1978). Optimum rehearsal patterns and name learning. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), Practical aspects of memory (pp. 625-632). London: Academic Press.
- Lavery, J.J. (1962). Retention of simple motor skills as a function of type of knowledge of results. Canadian Journal of Psychology, 16, 300-311.
- Lee, T.D. (1988). Transfer-appropriate processing: A framework for conceptualizing practice effects in motor learning. In O.G. Meijer & K. Roth (Eds.), *Complex motor behaviour: 'The' motor-action controversy* (pp. 201-215). Amsterdam: Elsevier Science.
- Lee, T.D., & Magill, R.A. (1983). The locus of contextual interference in motorskill acquisition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 730-746.
- Magill, R.A., & Hall, K.G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science*, 9, 241–289.
- Mannes, S.M., & Kintsch, W. (1987). Knowledge organization and text organization. Cognition and Instruction, 4, 91-115.

Melton, A.W. (1967). Repetition and retrieval from memory. Science, 158, 532.

- Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behavior, 16, 519–533.
- Nitsch, K.E. (1977). Structuring decontextualized forms of knowledge. Unpublished doctoral dissertation, Vanderbilt University, Nashville.
- Peterson, L.R., Hillner, K., & Saltzman, D. (1962). Time between pairings and short-term retention. Journal of Experimental Psychology, 64, 550-551.
- Rea, C.P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning*, 4, 11-18.
- Salmoni, A.W., Schmidt, R.A., & Walter, C.B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin*, 95, 355-386.
- Schmidt, R.A. (1975). A schema theory of discrete motor skill learning. Psychological Review, 82, 225-260.
- Schmidt, R.A. (1991a). Frequent augmented feedback can degrade learning: Evidence and interpretations. In G.E. Stelmach & J. Requin (Eds.), *Tutorials* in motor neuroscience (pp. 59-75). Dordrecht, The Netherlands: Kluwer.
- Schmidt, R.A. (1991b). Motor learning and performance: From principles to practice. Champaign, IL: Human Kinetics.
- Schmidt, R.A., & Bjork, R.A. (1989). New conceptualizations of practice: Common principles in three research paradigms suggest important new concepts for practice. Concept paper submitted to the U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.
- Schmidt, R.A., Lange, C.A., & Young, D.E. (1990). Optimizing summary knowledge of results for skill learning. Human Movement Science, 9, 325-348.
- Schmidt, R.A., Young, D.E., Swinnen, S., & Shapiro, D.C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 352-359.
- Schooler, L.J., & Anderson, J.R. (1990). The disruptive potential of immediate

feedback. In Proceedings of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

- Schulz, R.W., & Runquist, W.N. (1960). Learning and retention of paired adjectives as a function of percentage occurrence of response members. *Journal* of Experimental Psychology, 59, 409-413.
- of Experimental Psychology, 59, 409–413.
 Shapiro, D.C., & Schmidt, R.A. (1982). The schema theory: Recent evidence and developmental implications. In J.A.S. Kelso & J.E. Clark (Eds.), The development of movement control and co-ordination (pp. 113–150). New York: Wiley.
- Shea, C.H., Kohi, R., & Indermill, C. (1990). Contextual interference: Contributions of practice. Acta Psychologica, 73, 145-157.
- Shea, J.B., & Morgan, R.L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. Journal of Experimental Psychology: Human Learning and Memory, 5, 179-187.
- Shea, J.B., & Upton, G. (1976). The effects on skill acquisition of an interpolated motor short-term memory task during the KR-delay interval. *Journal of Motor Behavior*, 8, 277-281.

- Skinner, B.F. (1938). The behavior of organisms. New York: Appleton-Century-Crofts.
- Tolman, E.C. (1932). Purposive behavior of animals and men. New York: Century.
- Tulving, E., & Thomson, D.M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 87, 116-124.
- Winstein, C.J., & Schmidt, R.A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 677–691.
- Wulf, G., & Schmidt, R.A. (1988). Variability in practice: Facilitation in retention and transfer through schema formation or context effects? *Journal of Motor Behavior*, 20, 133–149.
- Wulf, G., & Schmidt, R.A. (1989). The learning of generalized motor programs: Reducing the relative frequency of knowledge of results enhances memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 748-757.

Forthcoming Articles

Commentary

When Predictions Create Reality: Judgments of Learning May Alter What They Are Intended To Assess Barbara A. Spellman and Robert A. Bjork

How Shall We Explain the Delayed-Judgment-of-Learning Effect? Thomas O. Nelson and John Dunlosky

Research Articles

The Contingency of Perceptual Processing: Context Modifies Equal- Loudness Relations Lawrence E. Marks

> Heuristics for Determining the Optimal Interval Between Checkups Roberta L. Klatzky, David M. Messick, and Judith Loftus

A Quantitative Genetic Analysis of Cognitive Abilities During the Second Half of the Life Span N.L. Pedersen, R. Plomin, J.R. Nesselroade, and G.E. McClearn

Research Reports

Perceptual Bias for Forward-Facing Motion Michael K. McBeath, Kazunori Morikawa, and Mary K. Kaiser

Violations of Monotonicity and Contextual Effects in Choice-Based Certainty Equivalents Michael H. Birnbaum

The Effect of Meaning Frequency on Processing Lexically Ambiguous Words: Evidence from Eye Fixations Sara C. Sereno, Jeremy M. Pacht, and Keith Rayner

Chronic Social Stress, Affiliation, and Cellular Immune Response in Nonhuman Primates Sheldon Cohen, Jay R. Kaplan, Joan E. Cunnick, Stephen B. Manuck, and Bruce S. Rabin

> Mathematics Anxiety and Science Careers among Able College Women Susan F. Chipman, David H. Krantz, and Rae Silver

> > The Disjunction Effect in Choice Under Uncertainty Amos Tversky and Eldar Shafir

> > > Genetic Influence on Risk of Divorce Matt McGue and David T. Lykken

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.