Art in Motion II
Motor Skills, Motivation, and Musical Practice

ADINA MORNELL (Ed.)

PETER LANG
Internationaler Verlag der Wissenschaften
Principles of Practice for Learning Motor Skills: Some Implications for Practice and Instruction in Music

This book’s goal is similar to that of the first book, based upon the original Art in Motion symposium held in Graz, Austria in 2008 (see Mornell, 2009) – the integration of (a) laboratory research in motor behavior and motor learning and (b) the field of music instruction and performance. It differs, however, by representing a subsequent step (beyond the 2009 book) in the development of these ideas. In this chapter, we have tried to apply some of the fundamental principles of motor learning – established in the research laboratories in kinesiology, physical education, experimental psychology, human factors/ergonomics, and related fields – to problems in high-level music instruction. These motor-learning experiments were done using mainly (unskilled) university undergraduates and artificial laboratory (non-musically-oriented) tasks. So, the question arises about the extent to which these principles could be applied to other real-world activities (e.g., high-level sports, music, and dance; speech-, occupational-, and physical-therapy, etc.) – high-level music included. We are not aware of any other attempts to integrate these two fields; so, hopefully, the present efforts to extend the outcomes of the 2008 Graz symposium will motivate others to continue in this direction. But, first, there are some major issues that must be addressed before we can be confident in applying laboratory-based motor-learning principles to instruction in high-level music.

Applying motor-learning research to music performance

First, we have the problem of measurement. Dr. Adina Mornell, our Symposium Organizer (in both Graz and Munich), has pointed out that, in high-level music, the audience absolutely expects the professionals to play or sing the correct notes flawlessly, with the correct pitch, and with the correct order and timing among them – no errors being displayed at all. When performers do make errors, these rarely disrupt the performance, that is, the errors are kept hidden from the eye and ear of the audience. In contrast, in high-level sports performances, errors are rather common. RAS’s former gymnastics coach at Cal Berkeley, Hal Frey, gave the following account of some of the action at the (American) National Collegiate Gymnastics Championships a few years ago: “Cal started on Horizontal Bar … Gymnast A, after his first several skills, lost his grip
on one hand and had to drop from the bar [suffering] an .8-deduction and probably his lowest score of the season. Following A was Gymnast B who [usually] does wonderful bar work, but he missed his regrasp on his release skill and suffered a .8-deduction ..." Clearly, high-level performances in music and gymnastics are considerably different; these examples from high-level gymnastics seem akin to having the pianist fall off the piano bench during the piece being played, which we don't see very often. Also, in the laboratory, where performances seem to be more similar to gymnastics than to music, we rely heavily on measures of error as a way of evaluating skill. We rely not only on errors in making the correct actions, but also the variability (a kind of error) in timing within the actions, and also the speed of the actions (where errors are commonly "traded-off" for speed), and so on. But, if high-level musicians do not even make audible or visible errors, it makes us wonder what we can learn about high-level music performance from studies of the production of errors in the laboratory.

A second, related consideration is that, in order to apply scientific methods and thinking to any field, one must be able to measure the phenomena of interest (here, musical performance). If what Dr. Mornell has said is correct (that high-level musicians rarely make errors - see Altenmüller in this volume), what should we measure in an experiment on musical performance in order to determine which among several practice variables has influenced learning the most? Also, when we mention this idea (i.e., that musicians rarely make errors), we often receive strange looks from people involved in high-level music. Obviously musicians do make errors, but the nature of these errors is perhaps different, or the errors are small and, hence, are not as evident as they are in simpler laboratory performances. If so, what are these errors, and how do we measure them? Or, if the measurement problem should amount to something other than the measurement of error, and includes evaluations of musical expression or artistic aspects, what are these exactly, and how then do we measure them? Clearly we must develop methods to measure performance in music in order to blend these two areas of study.

Here is another concern. The laboratory study of skill learning is usually done with essentially novel tasks, on which learners have not had very much practice. This is done chiefly to eliminate the confounding effects of the subjects' differential prior experiences with real-world tasks. As will be seen in the studies we present later here, often performers have
received only a few dozen practice attempts; it is rare to see laboratory investigations that involve more than a few hours' practice, although some examples do exist. On the other hand, as we all know, high-level musicians (as well as high-level athletes) have spent nearly countless hours practicing and honing their skills. Naturally, then, the question arises about whether or not the principles and theories developed in relatively unpracticed performers will apply to highly practiced musicians. Could there be another (separate) set of principles and methods that apply to very experienced people? We should point out that this concern has dampened the enthusiasm somewhat for the application of laboratory-motor-learning findings to training in elite athletes.

The last concern we mention is related to the nature of the tasks studied in laboratory vs. real-world tasks. Laboratory tasks are usually relatively simple, often involving only a single limb and/or using a performance that is relatively brief, although many exceptions to this notion exist. This is done, chiefly, so that performances can be recorded and measured with high degrees of precision and accuracy. However, most musical performances involve actions in many limbs in close coordination, and/or actions that unfold over many minutes. This raises the question of whether the laws, principles, and theories discovered in these simpler laboratory tasks would be expected to apply to the more complex musical tasks.

The final two concerns might not be so serious if we can come to some understanding (and agreement) about what to measure in musical performances, and how to measure it. Then, the last two concerns could be studied empirically using these measures to determine the answers. We suspect that, once the measurement issues are solved, studies to evaluate these latter questions can be done rather easily.

The next major section identifies and describes a number of findings — or principles — of motor learning that have been discovered in the laboratory. But first we consider the question of how motor learning is conceptualized by scientists in this area, as well as how it is measured.

Methodological and conceptual issues in learning

Learning is typically conceptualized as the relatively permanent acquired capability for performing some action, and it is analogous to a kind of
“accumulated quantity.” Importantly, the focus is NOT on the change in performance during practice (which can change in many ways that are only temporary); but rather our focus is on the change in this relatively permanent state that underlies performance. This concept is critical, and it forms the rationale for most of the methodology for studying motor learning in the laboratory.

Perhaps these ideas will be clearer with this analogy. Consider a glass of water that we “treat” by placing it in the freezer. After a while, the water changes its state, becoming frozen, solid ice. Now, if we take the “treatment” away by placing the glass on a kitchen counter, the water reverts to its original, fluid state. All of the changes caused by the “treatment” (applying sub-freezing temperatures) are completely reversible and temporary when the water is allowed to return to its starting temperature. Now, consider an egg, which we “treat” by placing it in boiling water. As a result of this “treatment” – analogous to what happened to the water when we applied the cold environment – the egg clearly changes its state in the boiling water; it becomes hard-boiled. Now, however, if we take the “treatment” away (i.e., remove the egg from the boiling water and allow it to return to its original temperature), the changed state of the egg remains; it is still hard-boiled.

What we saw happen with the egg, but not for the water, is the type of change we look for in studies of motor learning – changes that are relatively permanent – and we are not very much interested in changes that vanish when the “treatment” is removed. Put another way, in studies of motor learning, we are interested in finding changes that are “egg-like,” and we are relatively uninterested in changes that are “water-like.” But, how can we discriminate between these two situations empirically? Next, we describe a general method that has been used to answer this kind of question.

Consider this simple example. Let’s say that we have some piano task, on which the performance score decreases as performers improve (e.g., measures of error or time to complete the task). (We could just as easily have used a task for which the score increases with improvement [e.g., the number of correct actions]). Now, assume that we are interested in evaluating the effectiveness of two different methods of providing piano instruction – call them Piano Method A and Piano Method B. So, we conduct an experiment, having one randomly assigned group of subjects/
the change in ways that are relatively critical, and studying motor skills.

Consider a glass of water. While the water is relatively critical, the "treatment" water reverts to the "treatment" water and temperature. As the temperature to the water changes its state of the egg is like a glass of water. As we take a glass of water and change its state of the egg

is the type of changes that are relatively critical, and studies of "egg-like," "water-like." But, empirically? Next, consider this kind of piano task, skills improve (e.g., could just as easily provide piano providing piano B). So, we consider a group of subjects/subjects trained with Method A, and a second group of subjects being trained with Method B. If we measure the number of keyboard fingering-errors, plotting the average number of errors for each of the two groups, we could generate the curves shown to the left in Figure 1.

Given data like this from the music laboratory, typically, we would ask the question, "Which group learned more?" (or "Which method is more effective?"). At first glance, the answer seems nearly obvious: relative to the scores for Method A, the group with Method B showed (a) more accurate performance at the end of practice, (b) faster improvement (steeper slopes) during practice, and (c) more overall improvement across practice. But what if these differential gains shown by the A vs. B groups were just temporary (i.e., "water-like"), perhaps due to increased effort because Method B might be more interesting or more motivating,
or from less fatigue because Method B requires less energy expenditure? (There are a number of other possibilities, some of which we'll cover later in this chapter.) And given what we see in Figure 1, do we really know that Method B produced more (relatively permanent, "egg-like") learning than Method A? Scientists in this area would say, "No, we might suspect it to be so, but we are not sure."

So, with this general method, during the so-called "practice (or acquisition) phase," subjects practice under one of (in this example) two practice conditions (Method A or B). Then, during what is called the "test phase," subjects are assessed again but with the instructional conditions equated (and/or removed completely); see Schmidt and Bjork (1992), or Schmidt and Lee (2011). Further, this performance test is delayed in time from the end of the acquisition phase (e.g., from 5 minutes to several days or weeks). This delay allows (most of) the temporary ("water-like") effects of the treatment conditions (such as differential fatigue) to dissipate, allowing the relatively permanent ("egg-like") effects to remain. Then the groups are tested under equated conditions (such as removing the A and B methods for both groups) with either the same task (a so-called "retention test," here, the same piano-task used in acquisition), or with a different task (a so-called "transfer test;" here, this could be a test of accordion-keyboard performance). Testing performance under identical conditions for the two groups should ensure that any differential temporary effects will not reappear in the retention or transfer tests. Theoretically, at least, these tests provide a measure of the relatively permanent effects from the acquisition phase (with the temporary effects removed), allowing us to answer the question of which group learned more during the acquisition phase (where the two groups practiced under different conditions).

In our thought-experiment, we can imagine some alternative patterns of results that might occur; these are shown in the right side of Figure 1. Here, we show three possible results (call them Possible Retention [or Transfer] Outcomes 1, 2, and 3) that might be obtained in a transfer (or retention) test, done later on Day 2. (Of course, only one such outcome can occur in any given experiment.) Consider the first test result, Outcome 1. Here the A and B groups differ by the same amount as they did in the practice phase; we would say that all of the difference between the two groups in practice was due to relatively permanent effects, and none was attributable to temporary effects. Group B learned more than Group
Richard A. Schmidt & Timothy D. Lee

A. In the second result, the groups are the same in the transfer test, so that all of the differential gains of group B over group A in practice have dissipated by the time that they are tested. Now, we conclude that B and A learned the same amount. In the third outcome, the groups have reversed their positions, so that the group having had Piano Method A now shows less error than the group having had Piano Method B. In this case, we conclude that Group A learned more than Group B. There are, of course, many ways that these outcomes could be organized, and here we have shown just three of them. We'll see some others later in the chapter.

The crucial idea for understanding how to study learning is that we focus entirely on performance differences in the retention/transfer test, and pay almost no attention to any performance differences during the acquisition phase. (Actually, we would probably look at the acquisition-phase performances, but mainly to be informed about how these variables functioned during practice, and to get a more complete understanding of the results; but, we would not make any inferences about learning from them.) In the first example in Figure 1 (Retention Outcome 1), we would say that group B learned more than group A because - and only because - the B group outperformed the A group in the retention test. The same logic can be used for Outcomes 2 and 3.

These scientific methods are not only critical for studying learning in experiments, but they have important implications for the way we evaluate learning in the instructional process. As instructors, we use various techniques with learners during practice (e.g., we give encouragement, point out errors, give tips for performing differently, and so on). These things probably affect performance, at least somewhat; but we do not treat these resulting performance changes as being due necessarily to relatively permanent ("egg-like") gains (i.e., as learning). The critical information is how well the learner performs on some delayed criterion test (either retention or transfer) that has special interest. For example, in piano instruction this test might be the level of performance on next week's recital; for the gymnast it could be performance in the next meet; and so on. For all of these criterion performances, note that the various techniques the instructor uses are not available (i.e., they are withdrawn) for the criterion test. In this way, the retention/transfer test in research shares many similarities to the criterion performance in real-world teaching situations.
In the next section, we present some of the recent findings in the motor-learning literature. All of these findings were established using the methods just described. The focus is on some of the main principles that seem to make the biggest difference.

**Principles of practice organization**

The first of these principles is actually just a truism: There is little doubt that, in order to become a high-level/champion performer in music, sport, or one’s occupation, an enormous amount of practice time must be devoted to the task(s). Of course, this is no surprise to high-level musicians and athletes. Therefore, the important question becomes, “How does the learner (and/or the instructor) organize this massive amount of practice time, and/or what can be done during practice so that learning is maximized?”

**The specificity-of-training principle**

Here is another idea with great deal of intuitive appeal – even (or especially) for those not familiar with the research in learning: In order to maximize learning of a particular task under certain conditions, it is almost always most effective to match as much as possible the practice conditions with the (criterion) test conditions. For example, if you want to train your pianists to perform when fatigued, in front of an audience, and without any accompaniment, the optimal practice conditions would have them practice when fatigued in front of an audience and without any accompaniment. The evidence for specificity-of-training effects is so strong that many researchers consider it to be a general principle of learning – one that should be given primary consideration when an instructor considers various methods to organize practice. But, beyond this general principle, researchers have discovered many other important facts about the relative effectiveness of different methods for organizing practice. We begin our discussion of these with the description of an absolute “breakthrough” experiment by Shea and Morgan (1979), described next.

*Blocked vs. Random Practice*

Shea and Morgan (1979) considered the rather common situation where an instructor has a given amount of practice time, and has three different skills (say, Tasks A, B, and C) to teach in that time. One way to organize
the session, which seems to make good common sense, is to block the practice — that is, to split the time equally among the three tasks, practicing all trials of Task A in a row (i.e., in a concentrated “block” of trials), then all trials of B, and then all trials of C. Using this method, the learner can devote all his/her efforts to one task at a time, refine that performance, and not have to deal with interference from the other two tasks. Another way (there are countless other ways, of course) is to (pseudo-) randomize the practice — that is, practice one trial of Task A, then one of Task B, then one of Task C, ..., using a schedule in which no task is ever repeated on two consecutive trials and trials on all tasks are interleaved during the practice time. In experiments using this method the number of practice attempts (and, hence, the total practice time) is kept the same for the Blocked condition and the Random condition. Unlike the situation with Blocked practice, in Random practice, the learner would seem to be disadvantaged by the frequent switches between the different tasks.

Figure 2: Effects of blocked vs. random practice in performance vs. learning (Shea & Morgan, 1979). Blocked practice produces better performance than random practice during acquisition trials. But, random practice results in better retention performance, both immediate and delayed, and in both blocked and random retention orders, than blocked practice.
Using this basic design, Shea and Morgan (1978) studied the effectiveness of these two types of practice organization. They used tasks in which people made three-segment arm-movement patterns as rapidly as possible; the spatial pattern was different for three task-versions (A, B, and C). Let's see how these methods affected learning.

Figure 2 has the average time to perform these tasks, averaged across the three tasks, from the Shea-Morgan experiment. Look first at the performances at the end of the acquisition phase on the left. Data from all of the trial blocks during the acquisition phase have been removed from this graph, as the pattern of differences was similar to (or larger) than seen in the final block of trials. We might be tempted to conclude that Random practice (light bars) was detrimental for learning compared to Blocked practice (dark bars); the Blocked group improved more in acquisition, at a faster rate, and performed more rapidly at the end of practice, compared to the Random group. But recall our earlier discussion about how to evaluate learning, where we emphasized retention/transfer tests. Shea and Morgan brought their subjects back after 10 minutes (Immediate) and 10 days (Delayed), and tested them on the same three tasks, and also in either Random or Blocked retention orders. These results are shown in the right portions of Figure 2.

Consider first only the performance of the two groups that had the 10-minute retention test (Immediate, in the center of the figure). Regardless of the order under which the learners were tested (Blocked [Blo] or Random [Ran]), the group that had Blocked practice conditions in acquisition (dark bars) showed longer times (i.e., they were slower) in the retention test than did the group that had Random practice conditions in acquisition (light bars). Using our decision-rule described in the previous section, we must conclude that, for the immediate test of learning, these three tasks, Random practice in acquisition resulted in more learning than Blocked practice in acquisition.

Now consider the retention test done after 10 days (Delayed, on the far right of Figure 2). Again, regardless of the trial order in the retention test (Blo or Ran), the group having Random practice in acquisition (light bars), moved faster (i.e., they had shorter times) than the group with Blocked conditions in acquisition (dark bars). And, as with the results on the 10-minute test, the Random group was much faster than the Blocked
Richard A. Schmidt & Timothy D. Lee

effective-
tasks in which
rapidly as posi-
ions (A, B, and

eraged across
first at the per-
. Data from all
erminated from
or larger) than
clude that
g compared to
ing more in acqui-
ond of practice,
ussion about
transfer tests.
utes (Imme-
ne three tasks,
ese results are

s that had the
figure). Regard-
(Blocked [Blo] the conditions in
slower) in the
e conditions in
in the previous
earning, these
more learning

ed, on the far
the retention
acquisition (light
the group with
the results on
than the Blocked

This is a very surprising finding. Even though the Blocked group showed
smaller times (they moved faster) than the Random group during the
acquisition phase, Random practice produced more learning – and much
more learning on the Random retention test as compared to the Blocked
retention test, regardless of the time-delay of the retention tests (Imme-
diate or Delayed).

These organization-of-practice effects are extremely robust, and they
seem to operate in many different tasks and situations. For example, the
Blocked-vs.-Random effects occur in both motor learning and verbal
memory tasks; for children as well as older adults; for expert performers
of sports skills as well as novices, and in a wide variety of other situa-
tions. These effects are clearly not limited to simple laboratory tasks (see
Lee, in press, for a review). Because of this robustness, it probably has
importance for high-level music instruction as well.

What learning processes are operating in the studies? A dominant expla-
nation is this: Blocked practice allows the learner to do almost exactly
what he/she did on the previous trial, whereas Random practice forces
new retrieval and planning on each trial. With Blocked practice, the
learner can simply produce almost the same action time after time (with
a few minor changes, probably), but Random practice prevents this rep-
etition, and it requires the learner to abandon the action just made and to
retrieve from memory a different task, as well as to prepare it for execu-
tion. In this view, the facilitation in learning (for the Random group) is
based on the requirement to retrieve the action from memory and pre-
pare it for execution.

On the surface, this overall idea would seem to contradict the notion
of the need for high levels of repetition, mentioned earlier. However,
we don't think it does. The main point is that, while many practice tri-
als are critical for producing a high-level performer, this practice needs
to be done without rote repetitiveness. We see repetitious practice like
this often in real-world settings. The "classic" example is the golf driving
range, where players hit literally hundreds of identical shots in a row.
Players are often heard to say that their shots were wonderful in practice
(at the driving range), but that they could not perform the "same" shots
later in an actual golf game. What about the basketball player who practices 100 free-throws in a row, or the pianist who practices short parts of a piece over and over? How similar is this type of practice to competition or a recital?

Some other interesting phenomena are related to the notion of “spaced repetitions.” In many cognitive tasks, such as learning foreign-language vocabulary, scientists have shown that increasing the time interval (and/or the number of intervening trials) between repetitions of a given word degrades performance in practice on that word, but facilitates its performance at a retention test. This seems very much like blocked vs. random practice, to us. In addition, in what has been called “expanding retrieval practice,” the experimenter or teacher presents items using a schedule that increases systematically the spacing (either other items and/or time) between a given item and its repetition. As compared to presenting the same items with regular spacing, expanding retrieval practice degrades performance during practice, but facilitates retention (and, in our terms here, facilitates learning). See reviews of this work by Cepeda, Pashler, Vul, Wixted, and Rohrer (2006), Lee and Schmidt (2008), and Schmidt and Lee (2011).

However, these effects appear to be absent in long-duration tasks (Lee, 1981) that are mainly self-guided (such as the lab task called rotary-pursuit, or swimming, or steering a car). That these tasks appear to involve almost no advanced planning supports this view. The current thinking is that slow, lengthy, continuous tasks are driven mainly by the momentary conditions and sensory information, and they do not involve much advance planning. Even though music performances may be of long duration, they almost certainly involve planning of sub-sequences of notes. If so, we see no reason that would prevent these principles from being applicable to music instruction, for either high-level musicians or beginners. We point out some qualifications to this idea, next.

Random practice is not a “magic formula”

From the foregoing discussion, one might take the impression that random practice always facilitates the learning of all motor skills for all kinds of people in all situations. That is, random practice is some kind of “magic formula” that optimizes learning. It is not. In fact, we can point to some situations in which random practice is actually less effective for learning.
than blocked practice. In one example, Lee, Wishart, Cunningham, and Carnahan (1997) asked learners to perform a sequential button-pressing task; it involved a spatial-temporal pattern, whose timing could be represented via a series of timed "beeps" from a computer; Lee et al. termed this an (auditory) "model" of the action. When the model was presented just prior to a practice attempt, the auditory "beeps" specified the temporal structure of the action to be produced next. They used the usual Blocked- and Random-practice groups; but in a third group (also with Random practice) they augmented practice by providing this preliminary temporal information (the "model") just prior to every practice trial. Lee et al. found the typical random-blocked effects, where the Random-practice group performed with more error than the Blocked group during the acquisition (practice) trial blocks, but demonstrated less error in both immediate and delayed retention tests. But the results from the group with the "beeps" (Random + Model) were quite surprising. The modeled information produced very effective performance in the acquisition phase, and the performance of this Random + Model group was even slightly more accurate than the Blocked group. This is understandable, because the model provided very useful temporal cues for the learner immediately before each trial. But, in delayed retention, the performance of the Random + Model group had the largest error of all - even more error than did the Blocked condition. The modeled information seemed to obliterate any learning advantage provided by Random practice.

Scientists often discuss "retrieval practice" - the idea of practicing the act of retrieving the solution - as a way to strengthen memory. The idea here is that practice conditions that force the learner to practice the process of retrieval are beneficial when the learner must retrieve this answer in the future. The findings in the Lee et al. (1997) study are consistent with this idea. Random practice seems to require the learner to retrieve the temporal information before each trial, but providing the model presumably eliminated this requirement for the learners, as that information was now provided by the computer. The impact of this auditory cuing was sufficient to eliminate certain retrieval operations, and it ruined the effectiveness of Random practice completely.

These findings suggest that by giving advance information to learners, we might think that we are making the situation "easier" for them. We are, during practice; but in doing so, we have allowed the learners to avoid practicing retrieval operations that are critical for performance in
retention. Thus, random practice is no magic formula when other conditions serve to eliminate the very processes that make random practice effective.

In preparation for the 2008 *Art in Motion* conference and associated book chapter (Schmidt, 2009), RAS visited a rehearsal session of a symphony orchestra in a town near his home. At one point in the music, there was a period of silence (say, 5 sec. long), after which playing resumed; of course, all members of the orchestra were to resume playing at the same moment. In rehearsal, when this section of the piece was reached, the conductor/trainer would count aloud, “One-and-two-and-three-and-...” and all the musicians would start again on “5.” Later, RAS asked (a stupid question) whether this verbal information from the conductor was available during the actual concert performance; the answer was, “Of course not!” The conductor/trainer said that he was just trying to make it easier for the musicians to synchronize while in rehearsal. We are sure that he did make it easier. But, when viewed in terms of the results of Lee et al. (1997), he may have unwittingly blocked the musicians’ own internal time-keeping processes, perhaps even interfering with learning to synchronize on their own – something that would be required of them in concert.

Random-practice gains are not intuitive

One of the problems in using random practice in a learning environment is the resistance that might be encountered by both learners and instructors. The benefits of random practice are not intuitive, and indeed, are likely to be considered counterintuitive to traditional belief systems about how humans learn. Consider the findings of Simon and Bjork (2001). In their research, subjects were asked, throughout practice at a motor task, and also prior to a delayed retention test, to estimate how well they expected to perform on the future retention test. (These are called metacognitive judgments – i.e., knowledge about what you know.) Subjects in both the Blocked and Random groups estimated their future performances to be similar to the levels at which they had been performing in the acquisition trials. That is to say, the Blocked group predicted that their learning would be much stronger than that of the Random group; hence, Blocked subjects expected to produce much more effective retention performance than did the Random subjects.
When the subjects returned for their retention tests on the next day, Simon and Bjork once again asked them to predict their performance in retention. The finding from this test, illustrated in the left side of Figure 3, replicated what had been found on the previous day: the subjects in the Blocked group felt much more confident about their retention performance than the members of the Random group did. But, their actual results, shown on the right side of Figure 3, replicated what is typically found in studies of this type – Random practice produced more accurate performance in retention than Blocked practice.

![Figure 3: Effects of blocked vs. random practice on metacognitive judgments](https://example.com/figure3.png)

Two other points of interest illustrated in Figure 3 are important to consider. First, the level of retention performance predicted by the Blocked group was severely overestimated when compared to their actual performance in retention. Blocked practice leads the learners to think that they are learning well because of the success that they are enjoying in prac-
tice, when in fact they are not learning well at all. In effect, the learners are (erroneously) using their performances during the acquisition phase as predictors of future test performance. We wonder if such types of overestimation of "test" performance might account for why some musicians "choke" in concert (see Beilock, 2010). Perhaps their confidence from blocked practice leads to misattributions about learning, maybe curtailing the amount of practice undertaken, which is not revealed until too late (i.e., in the concert). The second point of interest concerns the relationship between predicted and actual levels of performance in the Random group. Instructors who consider using random practice in a learning environment might find that their learners become disenchanted with the practice environment (and perhaps, with the instructor, too). Random practice, or other methods that are deliberately designed to increase the overall "difficulty" of practice, have the potential to diminish confidence levels about how learning is progressing, which could reduce motivation to continue to practice as a result.

Thus, there is the potential that both types of practice - blocked and random - could result in reduced learning if these schedules influence the learners' confidence and motivation to practice. Over-confidence, resulting from blocked practice, might lead to an early cessation of practice. Under-confidence, resulting from Random practice, could lead to decreased motivation to continue to practice. Since the amount of practice is the primary factor leading to improvements in motor skill, both types of practice schedules, therefore, have potentially devastating influences if they are not used wisely. Fortunately, there are alternatives to the use of purely random or blocked practice schedules that retain the benefits of both schedules, but without the detrimental effects.

Alternatives to random practice

Conducting random practice involves several "difficulties." For example, is it logical for a guitarist to switch from song to song without ever repeating a section being practiced? Does it make sense for an orchestra to practice a piece just once before moving on to a new one? Does random practice become impractical when working with different instruments or in venues that require the learner to change rooms? Fortunately, researchers have anticipated and addressed some of these questions in experiments as well.
Richard A. Schmidt & Timothy D. Lee

Landin and Hebert (1997) examined issues like this in a study involving basketball shooting from different floor locations. They compared a traditional Blocked schedule, a schedule called “Serial” (e.g., A, B, C, A, B, C, A, ..., etc.), and a third schedule that could be considered a kind of “hybrid” approach to the first two. In this third schedule, the learner takes three consecutive shots at one location before moving to a new location, then takes three shots there, and so on. This type of practice could be considered to be \textit{moderate} in terms of task repetitiveness; it provided more repetitions (three shots) of a task before switching to a different task (compared to Random practice, which involves only one shot before switching); but this schedule did not exhaust all of the learners’ daily shots from one location before moving on to another spot (which occurs in blocked practice).

For example, is it ever repeatable for an orchestra to repeat the same music? Does random repetition instruments work? Fortunately, some important questions in the acquisition phase of learning are important. The results of this study are presented in Figure 4. Two sets of data are important here. Obviously, the data that appear on the right side of the graph, in which performance is compared in both Blocked and Serial...
retention tests, revealed that it was this Moderate group that produced the strongest learning effect. However, consider also the performance during the three practice days, on the left side of Figure 4. Performance in this portion of the experiment, too, was most successful for the Moderate group. Thus, performance during both the acquisition trials and the retention trials was facilitated. The Moderate group received some practice repetitions, but apparently not too many. When considered in light of the Simon and Bjork findings discussed earlier, in which attributions of learning are assigned on the basis of how one is performing during acquisition, this Moderate group probably would have benefitted in subjective judgments about their learning, too.

But one limitation of this approach becomes obvious when one considers the practical implications. How would one determine the amount of repetition to undergo during practice before switching to a new task version? There is arbitrariness when deciding how much blocked and random practice to combine in order to optimize performance, learning, and metacognitive judgments. Also, since there will be differences among individuals in learning rate, would it be prudent to advocate for just one type of hybrid schedule? Would skill-differences among people dictate that a unique hybrid schedule would be most appropriate for each person?

An experimental alternative to a fixed type of hybrid schedule (such as used by Landin & Hebert, 1997) is one that uses a rule-based algorithm to adapt the practice schedule to the performance capabilities of the learner. For example, subjects in an adaptive-learning study by Simon, Lee, and Cullen (2008) used a schedule in which “good performance” on a trial was always followed by a switch to a different task on the next trial, and “poor performance” was followed by a repetition of the same task on the next trial. Here, “good performance” was defined as an error that falls within a bandwidth of +/- 5% of being correct. This good-switch/poor-repeat algorithm determined the structure of practice for one group of individuals, and its effects on performance and learning were compared to two other groups of subjects undergoing traditional Random and Blocked practice. Thus, the first group’s practice schedule was adapted to the individual in order to suit the characteristic needs of each learner based on the rule-guided algorithm (see also Choi, Qi, Gordon, & Schweighofer, 2008). The findings of Simon et al. (2008) indicated that this good-switch/poor-repeat group resulted in performance
that produced performance
for the Moderate
trials and the
produced some
prac-
consideration in light
which attributions
performance
benefited in sub-
early considered, the amount of
new task ver-
le, learning, and
ences among
ate for just one
people dictate
schedule (such as
cessed algorithm
abilities of the
by Simon,
formance”
task on the
ompetition of the
was defined as
correct. This
ature of practice
and learning
ning traditional
scheme is
characteristic needs
also Choi, Qi,
(2008) indi-
performance
recently, Eliasz, Wishart, and Lee (2011) found that the opposite type of algorithm (i.e., good-repeat/poor-switch) to the one used by Simon et al. (2008) resulted in an even more successful adapted-learning schedule. The reasons why make good conceptual sense. For novice subjects learning novel tasks, the typical findings are large gains in improvement at the skill early in practice, and smaller gains later on. This general characteristic of the learning process was a critical factor in the results of this study. Eliasz et al. (2011) used a serial-movement task in which movement-time performance was reduced as the performance became more proficient. They chose a criterion for successful performance that was based on the subject’s previous history in the task (e.g., improving on one’s previous-best score). Due to the rapid gains in performance that almost always occur early in practice, many (or most) of the trials early in the acquisition process were a “success” (as defined by the criterion of improving on one’s best previous score). Learners quite often improved on their previous-best performance in the task in the early trials of learning a new skill. As proficiency improves, then bettering one’s previous-best score becomes more difficult to do. Thus, the pattern of task repetition and task switches for these two groups of algorithm-based practice schedules was very different. In the good-repeat/poor-switch group, practice on any one task was often repeated early in practice because of the relative frequency of “good” performances (improving on a previous best time). Later on, when it became more difficult to achieve the criterion, this algorithm resulted in much more frequent task-switches because the criterion was achieved much less often. In other words, this group (good-repeat/poor-switch) was characterized by an algorithm that resulted in mainly blocked practice early in acquisition trials, and mainly random practice later on. The opposite was true for the good-switch/poor-repeat group (used in Simon et al., 2008); their schedule was characterized as frequent random practice early in acquisition, and blocked practice later on.

This pattern of adapted practice resulting from the good-repeat/poor-switch schedule – that is, mainly-blocked- followed by mainly-random-practice – has been promoted as a type of practice schedule that makes
good pedagogical sense (Magill & Hall, 1990; Porter & Magill, 2010). Blocked practice during the early stages of learning allows the individual to focus attention and achieve a measure of success before task variation, in the form of random practice, is introduced. A good-repeat/poor-switch adaptive schedule that is based on the learner’s level of performance provides repetitive practice for someone who is slower at acquiring proficiency; but it is also capable of challenging the learner who progresses more rapidly. It is this concept of providing optimal challenges to learning to which we turn our attention next.

Challenging the learners during practice

Adapting practice-organization to meet the capabilities of the individual learners would appear to be a logical goal, both from a pedagogical and theoretical perspective. For example, Bjork (1998) has stressed the importance of introducing “desirable difficulties” into the learning environment – altering the learning environment in various ways such that the acquisition of knowledge involves some challenges to the learner. These difficulties are “desirable” to the extent that they (a) invoke processes that enhance the long-term retention of the information or skills, and (b) discourage processes that make the information subject to rapid forgetting. However, one might ask, “When does a ‘desirable difficulty’ become ‘undesirable’”? Also, “What is the effect on learning when the environment becomes too difficult?”

These types of questions motivated Guadagnoli and Lee (2004) to propose a research framework that developed the concept of “desirable difficulty” within the context of practice organization. The basic idea was that random practice increases the overall “difficulty” of practice – but whether or not it is a “desirable” feature to introduce into practice – to use Bjork’s concept – would depend on the nature of the task and the learner’s skill level. Tasks that are easy to acquire with practice would presumably benefit from the early introduction of random practice to increase the level of “desirable difficulty.” However, tasks that are complex and difficult to acquire would likely benefit less from early random practice because the high level of task difficulty might push the difficulty level to “undesirable.”

However, the concept of task difficulty is also modified by the learner’s skill level. A simple sequence on the piano, for example, might seem easy to the skilled pianist but very difficult for the less-skilled pianist. In other
words, task difficulty can only be defined as it relates to the skills of the person performing. Introducing random practice to adjust the level of "desirable difficulty" must be considered in the context of the task and the learner. Examples from research are presented next to illustrate these points.

Forehand- and backhand-tennis strokes were practiced by subjects of lower- and higher-skill in Hebert, Landin, and Solmon (1996), and these were practiced in either a blocked order or an alternating order (similar to random practice, in that the task is never repeated on two consecutive trials) over nine sessions. In two post-tests, the lower-skilled subjects who had practiced in a blocked order performed more skillfully than did the lower-skilled subjects who had practiced in a random order. In contrast, for the higher-skilled subjects, the post-tests were performed slightly more skillfully by the subjects who had undergone alternating practice than did those in blocked practice. These findings supported earlier laboratory findings by Del Rey, Wughalter, and Whitehurst (1982), demonstrating that the effects of the practice schedule were moderated by the proficiency of the learners. The findings have also been replicated in a golf task (Guadagnoli, Holcomb, & Weber, 1999).

Subjects of the same skill level vary also in terms of the effects of random and blocked practice, depending on the nature of tasks to be learned. Albaret and Thon (1998) asked novices to learn to reproduce patterns, originally seen on a video screen, by hand (without visual feedback). For any given subject in the experiment, the patterns were similar in complexity; all patterns involved either two, three, or four segments. Separate groups of subjects received either blocked or random practice of these two-, three-, and four-segment patterns, and later they performed retention and transfer tests (in the transfer tests, they made larger or smaller scaled versions of the learned patterns). Although random practice was generally advantageous for retention, the transfer results revealed that this was the case only for the two-, and three-segment patterns. No random-blocked differences were found on the four-segment pattern. Albaret and Thon (1998) suggested that the absence of random-blocked differences for the most complex four-segment pattern was due to the enhanced processing required to learn these patterns. In our terms, the most difficult pattern provided sufficient "desirable difficulty" to optimize learning, which was not facilitated further by randomly-ordered practice.
Choi, Qi, Gordon, and Schweighofer (2008) adapted practice to the skill levels of the learners (see Simon et al., 2008; Eliasz et al., 2011). Choi et al. asked their subjects to learn a manual-aiming task under different conditions of visuomotor transformation, and allocated a specific amount of allowable movement time to complete the task successfully. Practice was carried out over three sessions, during which the number of trials on a particular task was increased if it was needed (which gave more time to learn the task), or the allotment of time to complete the task was gradually decreased (presumably to increase the "desirable difficulty") for the higher-skilled learners. Retention tests at the beginning of each session provided estimates of how well the subjects were learning the task so that the necessary adaptations (more trials or increased difficulty) could be adjusted based on a learning measure rather than performance measures (as Simon et al. and Eliasz et al. had done). Adapting the number of trials and decreasing the time pressures to perform the task reduced error in retention dramatically, when compared to a fixed number of trials and fixed task difficulty in random-practice conditions.

Collectively, the studies discussed in this section serve to qualify, or limit, the effectiveness of purely random practice. For some subjects, random practice might hinder learning, at least until some initial gains in skill have been made. Some tasks may need to be practiced in a blocked order if their difficulty is inherently challenging. In general, recent research supports the concept that frequent monitoring of the individual's capability to perform the task, with appropriate increases (or decreases) in factors that influence "desirable difficulty" (random practice, task demands, etc.), have the potential to optimize the learning environment.

Some principles of augmented feedback

Augmented feedback is also of critical importance for learning. Augmented feedback is defined as: verbal (or verbalizable) information artificially added to the information the learner naturally receives from his/her performance (this natural information is called "intrinsic" or "inherent" feedback). It can be provided in a number of ways by an experimenter or instructor/coach. It has also been known as "knowledge of results" (abbreviated KR) if it describes the outcome in terms of the task's goal ("Your movement was too fast"), or "knowledge of performance" (KP) which is information about the movement pattern ("Your wrists should be higher above the keyboard"). It could be as simple as saying, "Right"
or "Wrong" or as complex as an audio/video replay of a person's performance compared with a video of a maestro's performance. Augmented feedback is critically important. If no other information is available about performance, then the performer may exhibit no learning at all from practice (Trowbridge & Cason, 1932).

Kernodle and Carlton (1992) asked learners to throw a lightweight, foam ball as far as they could with their non-dominant hand. Subjects who could not see where the ball landed, received augmented feedback in one of several forms: (a) as KR (the distance thrown), (b) a videotape replay of the throw (a kind of KP), (c) the video plus instructions about what aspect of the video to attend to (Cue), and (d) the video plus cues as to what to correct on the next trial (Transition).

All of the groups in the Kernodle and Carlton study increased their throwing distance with practice. But, there were sizeable differences among the groups. The group that received the video presentation plus instructions about what was important to view (Cue), and the group that received instruction about what to change on the next attempt (Transition), showed the most learning. Simply providing a video to watch without instruction (the KP group) provided little or no gain over simply telling the performer the distance thrown (KR group). Also provided were measures of throwing form which resulted in a pattern similar of results to that for throwing distance (Schmidt & Lee, 2011).

Many experimenters in the early- to mid-1900s studied the role of augmented feedback for learning in a wide variety of motor tasks. Much of this work was motivated by Thorndike's (1927) Law of Effect (see Adams, 1978, 1987, for reviews). This view held that a response followed by "reward" (feedback about success being "rewarding") tend to be repeated, while responses followed by "punishment" (e.g., error- or fault-information) tend not to be repeated. However, later in the 1900s, augmented feedback came to be regarded as information that could be used for subsequent changes in performance (see Adams, 1971). The general conclusion by about 1980 was that any variation of feedback that made it more immediate, more frequent, more accurate (or precise), more informationally "rich," or more "useful" would be beneficial for learning. We'll see in subsequent sections here that such a generalization is over-simplified.
Some augmented-feedback principles

Prior to about 1980, the prevailing view suggested that providing feedback after every practice attempt should be optimal for learning. According to this view, if a trial was performed without any augmented feedback, one might expect no learning to occur on that trial; or at least, learning should be minimal for that trial. More recent research does not support this idea (Salmoni, Schmidt, & Walter, 1984; Schmidt, 1991).

Feedback frequency

Winstein and Schmidt (1990, Exp. 2) gave augmented feedback on either every trial (100% feedback) or on half of the trials (50% feedback) over two practice days. The task was to produce an arm-movement pattern defined in space and time via a hand-held lever; it was scored in terms of root mean square (RMS) error – a measure of deviation of the movement’s pattern from the goal pattern, where smaller scores are more skilled. On trials on which feedback was given, it consisted of two forms of feedback: (a) a computer-generated trace of the subject’s attempt superimposed over the goal pattern (a form of KP), and (b) the computed RMS error (KR). After two sessions of practice with these feedback conditions, learners returned for retention tests (done without any augmented feedback) after five minutes (Immediate) or 24 hours (Delayed). These results are shown in Figure 5.

Giving feedback on 100% vs. 50% of the trials did not make very much difference during the two days of acquisition; from Figure 5 we can see that the performances on the last trial block in acquisition (left side) on each of these days are very close for the two groups, and slightly favor the 100% feedback group. But, on the immediate retention test, the 50% group outperformed the 100% group; and, this advantage was even greater on the delayed retention test. Clearly, the 50% group had learned more than the 100% group. These, and other findings like them, provide evidence against the previously held idea that more feedback is always more effective for learning. This has led a number of us to speculate about how feedback operates to generate such effects. We mention several of these ideas here; for more, see Schmidt and Lee (2011, Chapter 12).

One possible explanation is that highly frequent feedback acts as a kind of “crutch,” which guides the learner strongly to the correct action; but
this guidance does not generate the capacity to perform when the feedback is taken away — as it usually is in one's real-world criterion (test) performances. Under this view, which is sometimes referred to as the “guidance hypothesis,” the no-feedback trials tend to reduce this feedback-dependency.

A second possibility is that, because augmented feedback is so attractive to learners, it dominates the learner's attention immediately after the action, which interrupts the learner's attention toward the intrinsic feedback from the action (i.e., how the movement looked, felt, and sounded). Attending to augmented feedback presumably interferes with the learner's development of an error-detection capability. This is thought to be critical for detecting errors in the future and to correcting them — e.g., on a retention test.

Third, KR might induce the learner to change something on the next attempt — even if the action was already very nearly correct. Forcing changes on every trial might degrade the development of movement sta-
bility and consistency. These changes have been termed “maladaptive short-term corrections” by Bjork (1991; Schmidt & Bjork, 1992).

Fourth, augmented feedback might act to facilitate pre-movement information processing for subsequent trial(s) – in effect helping to “solve the problem” for the learner. If so, augmented feedback might act in a way analogous to blocked practice or to the way the auditory model interfered with learning in the study by Lee et al. (1997), discussed earlier (review Fig. 3 here), where the solution to the problem was made too accessible. In other words, augmented feedback might interfere with retrieval practice. Retrieval operations would be practiced to an increased extent when no feedback is provided on the previous trial.

Fifth, reducing the amount of feedback might increase the overall level of desirable difficulty in performing the task, in a manner analogous to the effects of random practice. In the absence of augmented feedback, the learner needs to generate subjective estimations of performance error in order to maintain levels of accuracy, or make improvements, and these require cognitive effort.

Which, if any, of these hypotheses might be correct must await further experimentation. Each of these hypotheses about these effects of augmented feedback are based on the idea that the learner is an efficient information-processor. As such, these hypotheses might not apply, or might apply in differently, when we consider learners with less-efficient processing systems (as with children, for example).

Consider the findings of Sullivan, Kantak, and Burtner (2008) and their study with child-subjects. Subjects from two age groups participated: young adults (range of 22-30 years) and children (range of 10-14 years). The task was similar to the one used in the Weinstein and Schmidt (1990) study. In fact, the filled squares in Figure 6, which plots the retention results of the adult subjects, replicated the Weinstein-Schmidt findings almost exactly. Retention performance was more skilled in the reduced-feedback group (which, here, had 62% KR) as compared to the 100% group.

But, contrast these findings with the results of the children, plotted as open circles in Figure 6. Using the same two feedback groups as were used with the adults, the children demonstrated a very different response
to reduced feedback; for the children the 100% group produced enhanced retention as compared with the reduced-feedback group. Instead of facilitating learning, as had been found with the adults (and by Winstein & Schmidt, 1990), learning was degraded in the children's reduced-feedback group.

![Figure 6: Effects of reduced relative frequency of augmented feedback in children vs. adults (Sullivan et al., 2008). Reducing the relative frequency of augmented feedback (62%) facilitated retention performance in adults (similar to Winstein & Schmidt, 1991), but degraded retention performance in children.](image)

What do these findings suggest about the hypotheses on the roles of frequent augmented feedback? One suggestion is that no-feedback trials do not work in the same way for children as they do for adults. Rather than the no-feedback trial acting as an opportunity to evaluate subjective-estimation processes, they might serve to reduce the children's attention or motivation to the task. Another idea is that feedback might need to be more frequent for children because they are less apt to use the information as a crutch to support performance. Still another suggestion is that the task had different levels of "desirable difficulty" for the adults vs. the children. Reducing feedback could have increased the "desirable difficulty" for adults, and promoted learning. In contrast, the
nature of the task was sufficiently difficult for the children, and reducing feedback might have elevated the "difficulty" to a level that was suboptimal, thereby degrading learning.

Summary Feedback

Another feedback variable that works similarly in many respects to the reduced-feedback effects seen in the previous section is called "summary feedback." In this research, the experimenter does not reduce the number of trials about which feedback is provided, but rather reduces the number of feedback presentations while at the same time increasing the amount of information provided. Here is how it works.

Say, for instance, that a teacher listens to a student play six different musical pieces in a dance suite. One way to provide feedback would be to give the student information about each piece directly after its completion. Another option would be for the teacher to listen to the student play all six pieces, and then give feedback about all six pieces at once (this is called summary feedback). These two feedback conditions both provide 100% feedback frequency. However, they differ in terms of the number of times the teacher intervenes to provide feedback (six times, or once). This is referred to as the "summary length," or the number of trials summarized – here one trial is summarized vs. six, respectively. Naturally, the amount of information provided during each feedback presentation will vary widely as well (i.e., information about one trial vs. information about six trials, respectively).

Research conducted in the 1960s by Lavery (1962; Lavery & Suddon, 1962), and extended years later by Schmidt, Young, Swinnen, and Shapiro (1989), revealed that the length of the summary (i.e., the number of trials summarized) had an important impact on learning. At least for relatively simple tasks, learning was facilitated when summary sizes increased. Several interpretations of these effects have been made. One is that increasing the summary length reduces the guidance properties of the feedback, as previously discussed. Another hypothesis is that the provision of longer summaries (but less often) has the effect of providing information about consistency (or variability) in performance, which might reduce the kinds of maladaptive short-term corrections that sometimes accompany every-trial feedback. A third hypothesis is that increased summary lengths, combined with increased information in these presentations,
increases the overall level of "difficulty" in the task. Given this latter possibility, it could well be the case that a particular summary length will be more or less effective depending on the nature of the task.

As a follow-up to their earlier study (Schmidt et al., 1989), subjects in Schmidt, Lange, and Young (1990) learned an anticipation-timing task, for which the nature of the task (anticipation timing) and the information provided as feedback (KP, instead of KR) were more demanding than in the earlier studies (i.e., in Schmidt et al., 1989). They used summary lengths of 1, 5, 10, or 15 trials, similar to their previous study. The retention outcomes however, were much different. Rather than learning being enhanced by increased summary sizes, Schmidt et al. (1990) found that a 5-trial summary was optimal for learning this task – and that 10- and 15-trial summaries were no more effective (actually, a little less effective) than the 1-trial feedback group.

Figure 7: Effects of summary feedback on novice and experienced learners as a function of task complexity (Guadagnoli et al., 1996). Larger summary sizes facilitate retention performance in experienced learners, regardless of task complexity. However, for novice learners, larger summary sizes facilitated retention only for the simple task; for a task of greater complexity, larger summary sizes degraded retention.
The findings of Schmidt et al. (1989, 1990) were replicated and extended in an experiment by Guadagnoli, Dornier, and Tandy (1996). Subjects in this study learned one of two force-production tasks—generating either unimanual or bimanual fist strikes to strike one or two padded force sensors. They were assigned to summary feedback groups that provided KR after 1, 5, or 15 trials. The retention results, summarized in left side of Figure 7, replicated the work of Schmidt et al. (1989, 1990) very well. For the simpler version of the task (unimanual), retention performance (i.e., learning) was enhanced as summary size increased from 1- to 5- to 15-trial summaries. But for the more complex version of the task (bimanual), retention performance was degraded as summary sizes increased. These findings fit very well with a desirable-difficulty interpretation, as the optimal overall level of difficulty created by the different summary-length conditions influenced learning differently depending on the inherent difficulty of the task.

However, Guadagnoli et al. took this experiment one step further. They repeated these three summary conditions, combined with the two task versions, in another six groups of individuals who had received practice on the tasks prior to data acquisition. In essence, the difficulty of both tasks was assumedly reduced for these more-experienced subjects, and hence the effects of the summary conditions would be expected to change as well. The findings for the experienced subjects are presented in right side of Figure 7. Notice now that the effects of the summary-length variable were similar for the simple and complex versions of the task—retention was facilitated as summary lengths increased.

Again, these findings converge well with desirable-difficulty concepts of learning. When a difficult task is attempted by novice learners, summary lengths need to be small to reduce the overall difficulty of the learning environment. However, as proficiency at the task increases, which effectively reduces the task’s overall inherent difficulty, then the environment needs to provide more challenge in order to promote learning. The level of difficulty must be raised in order to regain a level of difficulty that is necessary for effective learning.

Fundamental principles

To this point, we have presented several rather disparate and seemingly unrelated research findings (practice-scheduling and augmented-feed-
back effects]. There are several underlying and unifying principles (or themes) at work here, which we discuss next.

Retrieval practice

A critically important process in motor learning is retrieval practice which, we argue, seems to be an important basis of the random-blocked-practice effect. It is seen again in the research on feedback; feedback presentations that aid performance in practice (perhaps by making retrieval easier) were detrimental to learning as measured on long-term retention tests.

Creating desirable difficulties

Many of the practice variations that are the most effective for long-term retention (i.e., for learning) can be viewed as producing various difficulties for the learner during practice. Random practice made practice more difficult because the learner has to switch tasks on every trial, resulting in more effortful retrieval and planning for the next trial. Providing more intervening tasks or greater time, or both, (i.e., increasing spacing) makes performance more difficult, as the learner will experience some forgetting of the problem’s solution by the time it comes again. Feedback withdrawal on certain trials also makes performance more difficult for the learners, as they cannot be certain as to how well they are performing, and they are not guided to the correct action as strongly as they would have been with 100% feedback.

Creating appropriate challenges

Although methods of introducing desirable difficulties through such methods as random practice and reduced feedback frequencies seem to enhance learning in general, one must be cognizant of the learner’s capabilities and the inherent challenges of the task itself. When the task is sufficiently challenging on its own, one could argue against introducing methods that increase the inherent difficulty even further.

Final remarks and future directions

We began here by encouraging the recent attempts to bring various scientific principles and methods from the motor-learning area to bear on
instruction and performance in high-level music. We should not underestimate the challenges, however. We need to rationalize the differences between the motor-learning focus on slightly practiced, simple tasks, and music’s emphasis on very highly practiced, complex tasks. It could be that the kind of findings we have outlined here will be of considerable help in music – but we should be ready for challenges. At least, we have some scientific methods that will enable us to find out whether or not these principles are useful.

We also mentioned the challenge of measurement. If it is correct that high-level musicians rarely make errors, or the making of errors does not distinguish well between performers of different skill levels, then a focus on measures of error and similar phenomena in motor learning might not be very useful. Some other measurement scheme must be developed, perhaps one which focuses on various artistic and/or emotional aspects of performance that are scarcely touched in motor learning.

Next, we described some key features of the ways that scientists do experiments to study motor learning in the laboratory. This leads to two main outcomes. First, it provides fundamental concepts and basic methods that could be used in research in music and music instruction. Given the solution of the problem of measurement, these general methods could be used immediately in high-level music research. Second, given these methods in motor learning, we have described a number of research findings that seem to run counter to popular belief and/or to the common understanding of the learning process. These have encouraged the motor-learning field to re-think the principles, theories, and teaching methods that come from them. Perhaps these revised principles will be general enough that they will apply directly to music instruction. Perhaps they will not apply directly, and so they will have to be modified to account for the differences between performances in motor learning and high-level music. Now that we have the methods, and the desire, to accomplish these goals, it will be informative to watch future progress in this direction.
Would not underestimate the differences in difficult tasks, and easy tasks. It could be considerable of considerable least, we have whether or not

is correct that errors does not then a focus learning might not be developed, emotional aspects learning.

that scientists do This leads to concepts and basic music instruction. general methods research. Second, a number of belief and/or to have encouraged, theories, and revised principles music instruction. to be modified motor learning and the desire, to future progress in

References


Richard A. Schmidt & Timothy D. Lee


Note:
Some of the concepts and ideas presented here were also presented at the symposium *Art in Motion (AiM)* in Graz, Austria in 2008. See Schmidt (2009).