Ten guidelines for designing motor learning studies

RAJIV RANGANATHAN[1] | MEI-HUA LEE[1] | CHANDRAMOULI KRISHNAN[2]

[1] Department of Kinesiology, Michigan State University, East Lansing MI, USA.
[2] Department of Physical Medicine & Rehabilitation, University of Michigan, Ann Arbor MI, USA.

Correspondence to: Rajiv Ranganathan.
Associate Professor
Department of Kinesiology
Department of Mechanical Engineering
308 W Circle Dr Rm 126
East Lansing, MI USA 48824
email: rrangana@msu.edu
https://doi.org/10.20338/bjmb.v16i2.283

ABSTRACT

Motor learning is a central focus of several disciplines including kinesiology, neuroscience and rehabilitation. However, given the different traditions of these fields, this interdisciplinarity can be a challenge when trying to interpret evidence and claims from motor learning experiments. To address this issue, we offer a set of ten guidelines for designing motor learning experiments starting from task selection to data analysis, primarily from the viewpoint of running lab-based experiments. The guidelines are not intended to serve as rigid rules, but instead to raise awareness about key issues in motor learning. We believe that addressing these issues can increase the robustness of work in the field and its relevance to the real-world.

KEYWORDS: Motor learning | Skill | Experiments | Task | Analysis | Practice

INTRODUCTION

Motor learning is a central focus of several disciplines including psychology, kinesiology, neurophysiology, neuroscience, rehabilitation, and engineering. While this diversity of perspectives is a positive feature in terms of the development of new ideas and theories, it also brings associated challenges in terms of interpreting evidence and claims about motor learning. In our experience leading journal clubs, it is not uncommon to discuss a paper with a claim about "motor learning" in the title and end up questioning if the paper was really even about motor learning! A large part of this challenge is due to the fact that theoretical, conceptual, and methodological issues related to the design of motor learning experiments that are 'common knowledge' to researchers in one particular discipline may not always be accessible to researchers from other disciplines.

To address this issue, we provide a set of ten guidelines to raise awareness about these issues and navigate the design and analysis of motor learning experiments (Table 1). For each of these decision steps, we discuss common pitfalls and suggest recommendations, citing examples from both classic and recent studies of motor learning. Although several of these factors have been emphasized in prior work [1–5], the goal of this article is to synthesize this tacit knowledge to provide a step-by-step guide through the entire process from task selection to data analysis and interpretation. The paper is primarily intended for early-career researchers who are new to the field, but we hope that the issues raised can also serve as a starting point for discussions during interdisciplinary collaborations.

BJMB
Brazilian Journal of Motor Behavior

Table 1. Summary of the ten guidelines for motor learning studies. Each step in the process is listed with associated pitfalls and recommendations.

Steps	Pitfalls	Recommendations
Task Selection	• Lack of generalizability to real-world motor learning	• Explicitly define what the learning in the task entails. • Balance realism and insight
Instructions and Feedback	• Lack of clear instructions and feedback could affect strategy and outcomes	• Provide clear instructions • Provide feedback consistent with the instructions and maintain motivation
Practice Duration	• Practice duration affects the process being studied • Practice duration affects reliability and sensitivity	• Characterize a full learning curve for the task
Groups	• Two-group studies may not provide a complete picture of the learning effects and may result in misleading effect sizes	• Perform extensive initial characterization with multiple groups • Add realistic control groups
Sample Size	• Low sample sizes result in low power and unreliable estimation of effect sizes	• Provide sample size justification
Manipulation Checks	• Interpretations can be ambiguous without manipulation checks	• Provide appropriate manipulation checks
Tests of Learning	• Test conditions do not always align with learning in the task • Learning effects could be masked when viewing short snapshots of retention	• Ensure that test conditions are aligned with the definition of what the learning in the task involves • Increase the number of trials in the retention tests
Dependent Variable	• Choice between multiple dependent variables increases researcher degrees of freedom • Some dependent variables do not have desirable measurement characteristics	• Check if the dependent variable shows conceptual alignment, good measurement properties and mechanistic insight.
Measure of Learning	• Pre-tests can be problematic in some circumstances • Gain scores are not easily interpretable	• Minimize baseline imbalances by appropriate strategies • Exercise caution when baseline differences are inherent • Pre-register measures of learning and check for robustness of results with multiple measures
Processing Data	• Learning effects could be confounded by experimental artifacts (e.g., fatigue), processing artifacts, and statistical artifacts	• Add control experiments to reject experimental artifacts • Provide robustness checks for statistical artifacts • Provide open data and enhance transparency in visualization

STEP 0: DEFINING WHAT MOTOR LEARNING IS

One of the main barriers in motor learning research is that there is no universally accepted definition of motor learning across all disciplines and contexts [6,7]. Even at the behavioral level, definitions of learning have focused on several aspects including improvements in outcome, consistency, stability, persistence, adaptability, and automaticity [8]. Other definitions of motor learning have emphasized adaptation and reorganization of existing skills [9], changes in coordination dynamics [10], speed-accuracy tradeoffs [11], information pick up [12] and even decision making [13]. While there are certainly common characteristics across many definitions that fall under the classic view of learning as a "relatively permanent change in behavior" [4], it is important to note that the specific definition of motor learning adopted can have a major influence on many of the guidelines suggested here. For example, the question of how to measure learning can depend on whether learning is viewed as an improvement in the practiced skill (where learning would be characterized by a retention test with the same task goal and same practice conditions), an improvement in the adaptability or flexibility [14] (where learning would be characterized in terms of achieving the same task goal under different conditions), a general change in the movement capability (where learning would be characterized by a transfer test to examine the generality to other task goals that were not practiced), or a change in the underlying movement repertoire (where learning would be characterized by a 'scanning' paradigm examining the stability of different coordination patterns [15]). In this paper, we focus on general guidelines that we believe apply to many motor learning contexts, but these guidelines always have to be considered within the context of how motor learning is defined in that specific context.

STEP 1: TASK SELECTION

The task has a critical role in motor learning and is perhaps the biggest source of interdisciplinary differences. Given the criticism of 'applied research' for resulting in "disconnected pockets of data" that are unsuitable for the development of general scientific principles [16], the goal in lab-based settings has been to use somewhat artificial tasks to isolate specific aspects of motor learning (e.g., sequence learning, reducing variability). Therefore, choosing a task in lab settings needs to address two issues: (i) provide the feasibility to do experiments (e.g., tasks that can be learned in relatively short periods of time) and (ii) provide insight that is generalizable to real-world tasks.

Pitfalls

The use of 'simple' motor tasks can threaten generalizability to 'real-world' motor learning [1,5]. While it is not trivial to define apriori objective criteria on what makes a task simple, they can be roughly characterized as: (i) tasks where learning primarily involves figuring out the task goal and/or using existing movement capabilities to solve the task (rather than requiring a true change in the movement capability), and (ii) tasks with a lack of 'intrinsic information', which makes it difficult for a learner to judge their performance on their own, without feedback from the experimenter [17,18]. While it is likely difficult to identify these characteristics directly, they can be examined in the data, in which learning of such simple tasks is likely to be reflected as: (i) a sudden and very rapid improvement in

performance over a few trials of practice [5] with very little subsequent improvement (indicating that participants 'figured out' what the task goal was with no change required in movement capability), and (ii) a dramatic drop-off in performance when conditions are slightly changed (e.g., when performing the task after a delay or when withholding augmented feedback). For example, early work examining the role of knowledge of results has been criticized for using such tasks (e.g., draw a line of length 50 cm) where the learner does not have sufficient intrinsic sources of information, and therefore, has to rely on augmented feedback to even understand what the task goal is [18].

Recommendations

Explicitly define what the "learning" in the task entails. In view of improving our understanding of how studies from an experimental paradigm relate to others (including real-world learning), it is critical to answer two questions: (i) what is the problem that the learner has to solve to achieve good performance in the task, (ii) how much of this problem is known to the learner prior to learning (versus being discoverable only through practice)? For example, while most studies provide a 'score' that participants have to maximize or minimize, there may be important differences in how much participants 'know' about the task. In some cases, participants may know exactly how they need to improve this score (e.g., landing the ball closer to the target will result in higher scores). On the other hand, in typical reinforcement learning paradigms, the learner has no explicit knowledge of what results in higher scores and has to discover this relation through trial-and-error exploration. Making these types of distinctions explicit in the task description may be a critical step in separating out different types of learning studies, which potentially can lead to a better understanding of how findings from particular experimental paradigms generalize to other contexts.

Balance 'realism' and 'insight'. Given the inherent tradeoff between ecological validity and experimental control, it might be useful to consider tasks that resemble real-world contexts while also providing sufficient richness of measurement beyond just the task outcome [19]. For example, Haar and colleagues [20] devised a real-world pool task that measured the outcome (e.g., directional error), but also allowed for accurate measurement of whole-body kinematics and EEG. This is not to imply that 'more variables are better', but that well-justified dependent variables that go beyond the task outcome can provide insight into learning (also see Step 8). In addition, using tasks that allow for 'multiple pathways' to learn the task also allows insights into "how" learning occurred. For example, Sternad and colleagues devised a throwing task [21,22] where the main goal for participants is to reduce the variability of the task outcome, but the task was designed in a way that this improvement in variability could be achieved in multiple ways - decreasing noise, moving to more error-tolerant workspaces, or covarying movement parameters. The advantage of such tasks is that one can go beyond typical task outcome measures (where learning is almost certainly bound to have an effect) and look at the effects of learning at other levels of analysis (where the results may be more informative because it may not be easy to predict what the effects of learning may be) and the associated search strategies in the perceptual-motor workspace [23,24].

STEP 2: INSTRUCTIONS AND FEEDBACK

Once the task is defined, the next step is to decide on the instructions to the

participant and the associated feedback that is given to the participant. As mentioned in Step 1, one of the important aspects of the instruction is how they define the task goal for the participants. In addition to instructions, participants also typically receive some feedback during the task (usually provided using a score). In our view, while there has been an extensive body of literature on the effects of instructions or feedback – for e.g., manipulating attentional focus [25], given the powerful ways that instructions and feedback can channel learning, they should be considered carefully even in tasks where they are not the primary experimental manipulation.

Pitfalls

First, in the absence of clear unambiguous instructions, the learning problem may be "ill-defined" as participants may perceive the task goal differently from the way the experimenter intends it. For example, an instruction to "move as quickly as possible" could be interpreted as minimizing reaction time, minimizing movement time or minimizing the total response time.

Second, instructions or feedback may qualitatively change the strategy adopted to solve the task, especially when there are multiple, competing demands (e.g., "as fast and accurate as possible"). For example, in a speed-accuracy tradeoff experiment, even when other experimental parameters were closely matched, performance strategies under "time minimization" instructions (i.e., reaching to a target as quickly as possible) were qualitatively different from performance under temporal accuracy constraints (i.e., reaching a target as accurately as possible in a specified time) [26]. Similarly, feedback can also act to serve as a task constraint. For example, allowing for greater 'tolerance' in temporal precision changed the nature of the speed-accuracy function [27]. Given the recent emphasis on measuring speed-accuracy tradeoff functions as a measure of skill [11], these examples highlight the critical role of instructions and feedback in motor learning experiments.

Recommendations

Provide clear instructions. Clear instructions reduce uncertainty about the task goal and reduce between-subject variation in interpreting the tasks (which in turn reduces the need to eliminate 'outliers'). Adding a 'familiarization' block at the start of the experiment can also help provide a quick real-time check as to whether participants indeed understood the instructions.

Provide feedback consistent with the instructions. The feedback to participants should closely align with the instructions. This not only ensures that participants understand their performance but as seen later in point [8], this also allows for clearly defining dependent variables that can be used to track participants' compliance to the instructions. For example, in a steering task [28], the task instructions were to move a cursor from the start to finish as fast as possible, while staying within the boundaries of a channel. The feedback, therefore, involved a score that involved two terms - one that penalized the overall time and one that penalized the time outside the channel. By varying the weights of individual terms in an integrated score, it is possible to channel participants toward different strategies [29].

In addition to its informational role, feedback also has a motivational role [4] - so it may also be helpful to think of ways to keep participants engaged in the task over long

periods of time- e.g., using visual or sound effects to indicate good and bad performance [4], or providing a summary score at the end of a block of trials.

STEP 3: PRACTICE DURATION

Given the focus on 'simple' tasks, a large majority of motor learning studies tend to rely on a single session of practice, with possibly the addition of a 24-h retention test [4]. However, justifications for such practice durations are rarely specified explicitly. It is perhaps not surprising then that in addition to the type of tasks studied, these extremely short practice durations are another key barrier in translating findings from motor learning experiments to real-world learning or rehabilitation.

Pitfalls

First, from a theoretical standpoint, when the duration of practice is not clearly justified, it becomes unclear what process is being studied. Several theories have proposed that learning occurs in stages, with more 'cognitive' processes being engaged early on, followed by more automatic performance later in learning [16,30]. Second, from a measurement standpoint, the duration of practice impacts the reliability and sensitivity of the dependent variable. When practice durations are short, both within- [31] and between-subject [32] variability tends to be high, making measurements less reliable (although these factors depend to a large degree on the task and the stage of learning [33]). When practice durations are long, performance can reach a plateau, making the dependent variable less sensitive to detecting changes between different groups.

Recommendations

Characterize a full learning curve for the task. Rather than rely on preliminary pilot data, we suggest that initial experiments with a new paradigm involve learning with a relatively extended period of time so that the full learning curve can be characterized. The actual duration of this period will depend on the task, with the goal of trying to estimate when the performance reaches a relative plateau. For example, Reis et al. [11] examined the effect of brain stimulation on a speed accuracy task over 5 days, with retention periods up to 3 months. This type of initial characterization of tasks has three advantages - (i) it allows subsequent studies to potentially run shorter experiments depending on the research question, (ii) it can be helpful in determining if 'non-significant difference' between groups are due to performance plateaus, and (iii) it also allows for a more natural interpretation of effect sizes in terms of 'practice time saved'. For example, Day et al. [34] determined that abbreviated exposures to split-belt walking had the same effect as 4 days of practice at the task. This type of 'natural effect size' may be more intuitive in conveying the magnitude of effects for motor learning studies than relying on default standardized effect sizes like Cohen's d [35,36].

STEP 4: GROUPS

Most motor learning experiments that attempt to examine the effect of training paradigms rely on 'between-group' manipulations since it is reasonable to assume that motor learning cannot be 'washed out'. In this regard, motor learning studies have relied heavily on 'two group' comparisons - e.g., studies on variable practice typically have a 'constant' group and a 'variable group' [37]. However, defining these experimental groups is

critical in determining the insights that can be gained.

Pitfalls

The main pitfall of 'two-group' studies is that most manipulations involve a continuous variable that has non-monotonic effects on learning - i.e., there is an optimal parameter value or challenge point [38] at which learning effects are maximal, and the learning effects decrease on either side of this optimal value. Therefore, construction of two groups based on parameter values (e.g., selecting a low variability and high variability group) can yield different results depending on where the parameter values of these groups lie relative to the optimal value [2]. Second, even in cases where the groups are based on a variable that is categorical, two-group studies typically tend to be designed in a way to 'exaggerate' differences between groups [39], which can create misleading effect sizes and limit generalizability and real-world relevance.

Recommendations

Characterize 'dose-response curves' across multiple parameter values. Initial studies should characterize the dose-response across a reasonable range of parameter values to detect the presence of any optimal values. For example, an early study on the effect of distributed practice used five groups with inter-trial intervals ranging from 0 to 60 seconds and found no non-monotonic relation on performance or retention [40]. On the other hand, a study on summary feedback used multiple groups and found a non-monotonic change - i.e., both too frequent and too infrequent feedback had similar effects on learning [41]. While the actual research question is always critical in deciding how many groups are required, this type of extensive initial characterization with groups can help justify the choice of the number of groups (and the specific parameter values chosen for these groups) for subsequent studies using the same tasks.

Add realistic control groups. Even when a variable is clearly categorical, adding of 'realistic' control groups that represent reasonable choices for learning the task can provide greater clarity on the effect. A recent review characterizes in detail the different types of control groups that can be used in motor learning research [42]. One particularly important factor that adds a realistic control group in the context of motor learning is 'practice specificity' - i.e., when practice conditions match the test conditions [43,44]. For example, when examining the effects of variability on learning and consolidation, Wymbs et al. [45] used a number of control groups (including adding a 'practice specificity' group where no variability was added) to fully tease out the effect of adding variability. These types of data are not only useful for future studies, but also change the focus of motor learning studies from "whether" a particular type of practice matters to motor learning (which is almost always a "yes" in our view based on the idea that the 'nil' hypothesis of any intervention having zero effect is almost never true [46]) to "how much" it matters (which is much more informative).

STEP 5: SAMPLE SIZE

Sample size is one of the most important factors in experimental design. Most motor learning experiments tend to be small (typical sample sizes of 10-16/group) [2,47] likely indicating that this has become a heuristic determination rather than a justified determination. A difficulty with performing typical justifications with apriori power analysis is

that tasks tend to vary between studies [2], making it difficult to extract effect sizes from prior work.

Pitfalls

The pitfalls of low sample sizes have been extensively covered in other literature [48,49] - so we just highlight two main points: (i) low sample sizes have low power, which can lead to missing true effects, and (ii) a literature with low powered studies means that published effects will tend to have inflated effect sizes [48].

Recommendations

Justify sample size. Sample size justification statements are critical for judging the informational value of a study. Other than apriori power analysis, these can include other types of justifications such as accuracy or resource constraints [50]. McKay et al. [51] provide an example of sample size justification where the sample size was based on a prior study, with adjustments to the p-value for sequential analysis. In light of steps 3 and 4 that raise the need for initial characterization of long-term studies with multiple groups, it is critical that these studies also have sufficient sample sizes for reliable estimation of effect sizes.

STEP 6: MANIPULATION CHECKS

Manipulation checks refer to tests that examine if the designed intervention had the desired effect on the participant. While certain types of interventions do not require such checks since they are directly under the control of the experimenter (e.g., the inter-trial interval), other variables are designed to elicit specific responses during practice (for e.g., increasing motor variability using force perturbations). Often, with the emphasis that practice during training may not be reflective of 'true' learning (i.e., the learning-performance distinction) [52], there is sometimes a tendency in motor learning studies to only focus on the end-result without interpreting if an intervention had the desired effect 'during' practice.

Pitfalls

Without a manipulation check, interpretations of results can be ambiguous. For example, a study on contextual interference (typically manipulated using a blocked vs. random schedule) needs a manipulation check that the random schedule indeed created more interference during practice. This manipulation check is usually done by examining if there were higher errors during practice for the random group. Without this check, it is difficult to disambiguate a true 'null result' (i.e., contextual interference did not have a significant influence on learning) from a 'failed manipulation' (i.e., the experimenter's attempt at introducing contextual interference was not successful).

Recommendations

Provide manipulation checks. Manipulation checks can vary depending on the type of study. In some cases, manipulation checks can be done using variables that are already directly recorded during data collection. For example, Cardis et al. [53] showed perturbations of the null and task space increased variability in these dimensions. In other cases, there may be the need for separate manipulation checks - e.g., Grand et al. [54] used

both the intrinsic motivation inventory and electroencephalography to examine if the manipulation resulted in increased motivation. When manipulation checks 'fail', it might be critical to consider all aspects of experimental design mentioned in Steps 1-5 (the instructions to the participant, the amount of practice, etc.) as well as the measurement resolution of the variable being used for the manipulation check. However, one caveat when introducing these manipulation checks is to also minimize the possibility that the manipulation checks themselves change the behavior of the participant [55].

STEP 7: TESTS OF LEARNING

Perhaps one of the most important insights in motor learning is the 'learning-performance' distinction - i.e., that changes in performance during practice are not always indicative of learning [3,52]. This insight has resulted in the use of separate 'tests' of learning-e.g., delayed retention tests or transfer tests as conditions for measuring 'true' learning. This raises a critical question as to what test is most applicable for measuring learning.

Pitfalls

First, there has been an overreliance on the "24h No-KR test" as the 'only' acceptable condition for measuring learning [17]. While the 24h No-KR condition can be a good indicator of learning in some cases, both the "24h" and the "No-KR" portions of the test can sometimes be problematic. First, the 24h period is intended for dissipating temporary effects, but whether this makes it a 'true' test of learning may depend to a large degree on the task and the timescale of learning involved. For example, rehabilitation paradigms that attempt to change movement capacity over several weeks typically involve follow-up tests in timescales of weeks or months. Second, the No-KR part of the test is intended to provide a 'stable' performance by minimizing possibilities for learning "during" the test. However, in many cases, because information about performance is available naturally through sensory information outside of "augmented feedback" provided by the experimenter, trying to prevent learning during the test has led to artificial measures such as 'blindfolding' the individual and/or using headphones playing white noise. In such cases, the dramatic change in feedback information may induce the learner to adopt qualitatively different strategies during the test, which are unrelated to the original learning.

Second, these tests of learning are typically designed to measure only a 'snapshot' of the process (usually as short as 10 trials). This means that the snapshot itself can be unrepresentative of true learning because it can include temporary effects such as warm-up decrement, especially when administered after a time delay, or when there is a sudden change in the context [17].

Recommendations

Justify test conditions. We recommend that the test conditions are aligned with the definition of what the learning in the task involves (mentioned in Step 1) and what aspects of information are considered 'part of the task'. For example, Lokesh and Ranganathan [28] used a haptic feedback study where the test involved removal of haptic feedback, but retained concurrent visual information of the cursor because this visual information was considered part of the task. Similarly, the timeline of such a test can be aligned with respect to the total learning time and other task factors such as fatigue. For example, a 24 h retention test may be justifiable for tasks with a one-day practice duration,

but longer timescales of learning may potentially require longer time intervals. For example, in a balancing task that was practiced over 6 weeks, the retention test was made 3 months after the last practice block [56].

Make tests more than a single 'snapshot'. Including a retention test with a relatively large number of trials can provide the ability to separate out temporary and relatively permanent effects. For example, studies of massed and distributed practice show that while there are effects in the first few trials of the next day, these differences are quickly eliminated within the next few trials of practice [57]. A further alternative is to consider a series of retention periods - for example, Reis et al. [11] measured retention at 5 time points from 3 days until approximately 3 months after the last practice. This type of data could characterize retention over time, which is critical for studies that examine learning in the context of a change in the movement capability. In addition to testing over time, transfer tests [1,58,59], which involve testing in conditions that were not originally practiced, can also provide a richer description of learning. However, a critical piece in transfer tests is determining the appropriate transfer conditions so that it is clear how these transfer conditions relate to the originally practiced task. van Rossum [60] emphasizes the use of task analysis as a means of designing transfer tests so that the results from transfer tests can yield useful information about what was learned. For example, in sequential timing tasks, transfer tests with the same or different relative timing patterns as those during practice were used as a critical test to distinguish whether the benefits of variable practice were due to the formation of motor schemata or contextual interference effects [61,62].

STEP 8: DEPENDENT VARIABLES

Dependent variables in motor learning studies have to accomplish two main goals - (i) the variable sufficiently captures the improvement in task performance, and (ii) it provides 'insight' into how improvements are occurring. Part of the challenge in defining these variables is that motor performance is multidimensional and there is usually no unambiguous choice of variables. For example, in a golf putting task, the dependent variables used to measure learning could be (i) the percent of putts made, (ii) the number of points scored in a scheme defined based on the distance from the hole (e.g., 10 points for making the putt), (iii) the absolute error in terms of the distance from the target, or (iv) the variable error in terms of the consistency of the putts. In some cases, the dependent variable may even be derived 'post hoc' at the data analysis stage - for e.g., the number of putts that were within 1 inch of the hole, or a score that weights the absolute error in the putting task by the time taken to prepare for the putt.

Pitfalls

First, the choice between several dependent variables (especially when they are derived post hoc) can inflate researcher degrees of freedom [63]. For example, in the golf putting task described above, the choice between many possible dependent variables can result in a situation where some variables show statistically significant differences but others do not. If undisclosed, such flexibility in the choice of dependent variables can lead to a large increase in the false positive rate [63].

Second, some dependent variables do not have desirable characteristics either from a measurement standpoint or a mechanistic standpoint. For example, from a measurement standpoint, measures such as number of putts made can be insensitive to

changes in the magnitude of errors. In terms of mechanistic insight, some dependent variables may not provide sufficient information about the performance. For example, in 2D aiming tasks, using only a scoring system (or radial error) may mask important changes in terms of changes in bias or variability [64,65].

Recommendations

Check if the dependent variable shows conceptual alignment. From a conceptual standpoint, the dependent variable should closely align with the instructions and feedback to the participant. As Newell [66] states "it seems unreasonable to evaluate the effect of an independent variable primarily through criteria different from that originally stressed to the subject" Moreover, alignment with instructions and feedback reduces researcher degrees of freedom by minimizing the potential for 'post hoc' derived dependent variables. For example, in a gait tracking task which involved the participants to match a target template with their foot, Krishnan et al. [67] define the error measure only in the spatial dimension using an area metric (as opposed to say using RMSE, which would need to involve assuming something about the temporal component that were not part of the instructions). In cases where the dependent variable may not be apparent immediately from the instructions and is a combination of multiple variables (say in a task that emphasizes both speed and accuracy), it might be ideal to pre-register this analysis so that there is transparency about flexibility in data analysis [63].

Check if the dependent variable shows good measurement properties. From a statistical standpoint, dependent variables should be sensitive enough to track changes with learning. Although measures such as error rates or % success may be used in some contexts because they have functional significance (i.e., in a real game, "near-misses" do not count), in general they may be too coarse-grained to detect differences. While it is desirable to have these dependent variables in real-world units (e.g., error measured in cm), this may not always be feasible and may require other approaches. For example, studies on the free-throw use a point-based system (e.g., 5 points for a swish) to improve the sensitivity of the dependent variable [68]. However, it is important to note that any system developed should also be considered on a conceptual basis - i.e., if participants get scored higher for a free throw that goes directly in the hoop compared to one that goes in after bouncing off the rim or the backboard, this information should also be communicated to the participant through instructions.

Check if the dependent variable provides mechanistic insight. Finally, consider dependent variables that yield 'insight' into the underlying process of motor learning. In many cases, these may be considered 'secondary' variables that help delineate different processes. For example, the dependent variable in many aiming tasks is an absolute error (AE), but this measure is a combination of constant error (CE) and variable error (VE) [69]. Therefore, while the AE can be the 'primary' dependent variable (based on conceptual grounds that it was the instruction to the participant) [66], breaking down the AE into secondary variables CE and VE can provide additional insight into 'how' participants got better. Similarly, Lee et al. [70] use movement time as the primary dependent variable in a cursor control task, but also measure the path length to examine if increased movement times are due to changes in movement speed (which would not change path length) or taking more circuitous paths (which would result in increased path length).

STEP 9: MEASURE OF LEARNING

The measure of learning used in the analysis is extremely critical, as they form the basis for inferences drawn from the experiment. Similar to the debate about measuring learning in many other fields, there are two distinct philosophies for measuring motor learning - proficiency (i.e., how good is performance after practice) and growth (i.e., how much has performance changed after practice), and a multitude of measures have been used to quantify learning based on both approaches [71]. Proficiency-based measures typically rely on the absolute performance level at the end of practice (e.g., on a retention test) to measure learning. On the other hand, growth-based measures typically use a pre-post experimental design, where the change between initial and final performance levels are compared across groups. This change can be (i) a gain score (i.e., final performance - initial performance), (ii) a gain score that is 'normalized' in some way (e.g., final performance/initial performance *100) or (iii) a post-test score or gain score that uses the pre-test score as a covariate. In addition, the 'rate' of learning is often also computed as a way to capture if one group learns faster than the other. As in the case with the choice of dependent variable, the presence of multiple measures to assess learning can create a challenge, as it increases researcher degrees of freedom.

Pitfalls

First, the use of 'pre-tests' in motor learning has been criticized on two grounds [4] - (i) that pre-test scores in motor learning are unreliable because early trials at a task generally are extremely variable and show very poor correlation with final performance, and (ii) extending the trials in the pre-test to get a reliable measure of baseline performance essentially provides practice at the task, thereby minimizing the room for improvement during the actual intervention.

Second, the use of gain scores has been criticized as a learning measure [72] especially in cases where the pre-test scores are not the same across groups. For example, since the gain scores are skewed by the initial performance level, individuals with a lower pre-test scores could appear to have greater 'gains' even if their final performance was lower or similar to an individual with higher pre-test scores [71]. This will especially be the case in situations where the performance approaches a plateau. While normalizing the gain score (e.g., expressed as a %) can offset some of these issues, this also has to be treated with caution since the normalization procedure makes assumptions about the form of the learning curve.

The problem with a gain score is even more obvious when using a 'relative retention' measure (i.e., a difference score computed between the last block of practice and the retention test). In this case, the magnitude of relative retention is heavily dependent on performance during practice—a measure that may be affected by factors other than learning [71]. Moreover, if groups underwent different types of intervention during practice, the performance at the end of practice is contaminated by the effects of the intervention itself, making it an unsuitable measure to compare learning in the two groups.

Finally, some motor learning studies use curve fitting (e.g., fitting an exponential) as a means to quantify 'rates' of learning independent of performance level. While this argument is true 'in theory', this relies on two major assumptions - (i) the form of the function actually matches the form of the learning curve (i.e., if an exponential fit is made, that the curve is actually exponential), and (ii) the data have very minimal noise, and the

function fits these data extremely well at the individual level. Without full knowledge of these assumptions, it is difficult to interpret results from curve fitting.

Recommendations

Minimize baseline imbalances. When initial performances at baseline (i.e., the pre-test) are similar between groups, the type of the dependent variable used in the analysis has a minimal effect on the outcome - i.e., the use of a change score or the final score will lead to similar conclusions. Hence, it is critical to ensure that adequate attempts are made to minimize baseline imbalances. This can be done using three strategies - (i) increasing sample sizes: small sample sizes are more likely to create baseline imbalances; hence, we recommend using adequate sample sizes to ensure that the baseline performance reflects the true group means [73,74], (ii) stratifying groups based on initial performance including treatment of outliers [75], and (iii) using an analysis of covariance (ANCOVA) to adjust for any residual differences in baseline for estimating an unbiased intervention effect [76,77]. In cases where 'pre-test' measurements are either not feasible or consistent, it might be helpful just to rely on larger sample sizes and use the final test performance as the measure of learning.

Interpret results with considerable caution when baseline performances are different. When baseline imbalances are inherent to the research question - e.g., when comparing age-related differences [78,79] or the effects of neurological injury [80], the research question can be 'ill-defined' to some degree and the results have to be treated with considerable caution since statistical techniques like ANCOVAs that account for baseline differences can yield misleading results [81]. Instead, it might be fruitful to think about potential experiments where ambiguity can be resolved. For example, if a particular study shows that rates of learning in children are higher than adults (but is ambiguous because children still perform lower in an 'absolute' sense relative to the adults), then by extending the practice period, one can test in a follow-up experiment whether there is a point where the children 'outperform' the adults even in terms of absolute levels of performance. This result would help resolve the ambiguity since both absolute and rate measures would show that the children outperformed the adults.

Pre-register learning measures and robustness checks. Again, as with the choice of the dependent variable, an important way in improving transparency is to pre-register the learning measures of the study and make data openly available to other researchers. In particular, adding robustness checks (e.g., comparing learning using multiple learning measures) to check if the outcomes change with the type of learning measures will help improve the interpretation of the study results.

STEP 10: PROCESSING DATA

Given the fact that most motor learning data inherently involve change in performance over time (i.e., there is no 'stable' performance), with significant within- and between-subject variability, even seemingly standard data analysis procedures can sometimes result in 'artifacts' that can be misleading. In addition, the effect of non-learning related factors on performance (such as fatigue) can also pose challenges during data analysis. Therefore, it is important to examine the effects of steps in the data analysis pipeline.

Pitfalls

We highlight three examples of artifacts arising in motor learning studies at different levels within the data analysis pipeline:

Experimental artifacts. While the general 'learning - performance distinction' is well recognized in the context of needing retention or transfer tests [3], many questions in motor learning require direct analysis of performance curves - e.g. the study of 'online' and 'offline' learning that examine changes in performance 'during' and 'between' practice sessions or even at the level of trials [82–85]. One phenomenon observed in these experiments is 'offline consolidation' (also referred to as 'reminiscence' [86]), where there is a seemingly distinct improvement in performance after a rest break. However, there is a potential for performance artifacts (e.g., effects of inhibition or fatigue) in these types of analyses [57,87] and it is worth noting that these consolidation effects are typically observed in speeded tasks (e.g., produce as many typing sequences in a given time period, or tracking moving targets continuously).

Processing artifacts. General data processing procedures involve some type of averaging to reduce the 'noise' in the data. However, relying on "averaged" data across individuals can substantially alter some types of inferences. From a theoretical point, one of the most well-known artifacts due to averaging is the form of the learning curve which can differ between power-law and exponentials depending on averaging [88–90]. As a result, there is potential for artifacts when estimating individual parameters based on group-averaged data. A second type of averaging that is usually done across trials within a block can mask temporary effects like warm-up decrement [17,89] and create the illusion of a 'discontinuity' in the learning curve even when the learning curve is continuous [87].

Statistical artifacts. In the analysis of individual differences, 'mathematical coupling' occurs when the response variable directly or indirectly contains all or a part of the predictor variable [91]. For instance, studies evaluating initial performance to predict 'gain scores' are expected to see significant associations simply due to the fact that the gain score depends on the initial performance. Relatedly, methods that involve creating 'post hoc' groups based on dichotomizing a continuous response (e.g., analyzing the difference between high and low responders based on a median split) have been criticized from a statistical viewpoint as creating arbitrary or illusory distinctions [92]. It is especially important to note that standard procedures used in other domains (such as repeated exposures to examine if the 'response' is reliable within an individual) are not generally applicable in motor learning since it is not possible to wash out prior learning. As a result, results from these types of analyses must be treated with even more caution.

Recommendations

Control experiments for experimental artifacts. In cases where the analysis is based on performance curves, the use of control experiments that can disambiguate temporary effects of performance from learning are critical. For example, to examine if offline learning is truly distinct from 'recovery from inhibition/ fatigue' effects, it might be useful to manipulate the practice/rest interval or perturb neural activity during rest period [93].

Robustness checks for processing and statistical artifacts. The impact of specific procedures can also be examined by showing how results are impacted by changes in specific parameter choices. For example, when using a parameter based on a group-averaged curve fit, Reis et al. [11] show sensitivity analyses to changes in this

parameter. Using simulated data with known properties can also be a critical tool to check the impact of a particular processing procedure. For example, Smeets and Louw [94] showed how the decomposition of variability can be sensitive to the choice of variables used in defining the task.

Providing transparent visualization and open data. Finally, it may be difficult (if not impossible) for one paper to identify all possible artifacts and/or perform all possible robustness checks. Therefore, providing transparent visualizations that go beyond simple bar graphs (e.g., showing individual performance curves with minimal or no averaging) [95] and using open science practices like publicly sharing analysis and data can be extremely critical in improving the quality of motor learning science [96].

CONCLUSION

Overall, the pitfalls and recommendations highlight two broad themes in motor learning that require attention. The first theme relates to the relevance of motor learning studies to the real world. As highlighted earlier, while we agree that definitions of motor learning will vary depending on context and discipline, it is perhaps also important to take a pragmatic perspective that in some way, the ultimate goal of motor learning experiments is to be able to apply this knowledge to the real world. In this regard, while broad 'principles' of motor learning are often mentioned in the context of fields like rehabilitation [6,97], it is difficult to gauge the actual impact of most current motor learning paradigms on these fields. For example, a recent review of stroke rehabilitation literature found that only 8% of studies even mention 'basic science' studies (motor learning experiments in animals or humans) in the Introduction [98]. We hope that by raising awareness of several issues that hamper real-world relevance (choice of task, length of practice duration, etc.), the guidelines spur researchers to move outside 'traditional' paradigms in their own subfield with the goal of increasing relevance.

A second theme that emerges from the guidelines is the need for initiatives that are not just at the level of a single investigator or a lab but at the level of a whole research community. As highlighted in the context of developing 'model task paradigms [2], many of the proposed recommendations (larger sample sizes, more groups, increasing practice duration, preregistration, sharing open data) require a greater investment of time and effort compared to current publication practices. As a result, we hope that the guidelines spur discussion not only about larger-scale collaborative efforts, but also the need for recognition of such efforts at other levels such as by journal editorial boards, hiring, and promotion and tenure committees.

REFERENCES

1. Newell KM. Motor skill acquisition. *Annu Rev Psychol*. 1991;42:213-237. doi:10.1146/annurev.ps.42.020191.001241
2. Ranganathan R, Tomlinson AD, Lokesh R, Lin TH, Patel P. A tale of too many tasks: task fragmentation in motor learning and a call for model task paradigms. *Exp Brain Res*. 2021;239(1):1-19. doi:10.1007/s00221-020-05908-6

3. Salmoni AW, Schmidt RA, Walter CB. Knowledge of results and motor learning: A review and critical reappraisal. *Psychol Bull*. 1984;95(3):355-386. doi:10.1037/0033-2909.95.3.355

4. Schmidt RA, Lee TD. *Motor Control and Learning*. 5th ed. Human Kinetics; 2011.

5. Wulf G, Shea CH. Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychon Bull Rev*. 2002;9(2):185-211. doi:10.3758/bf03196276

6. Krakauer JW. Motor learning: its relevance to stroke recovery and neurorehabilitation. *Curr Opin Neurol*. 2006;19(1):84-90.

7. Newell KM, Ranganathan R. Some contemporary issues in motor learning. In: Sternad D, ed. *Progress in Motor Control: A Multidisciplinary Perspective*. Advances in Experimental Medicine and Biology. Springer US; 2009:395-404. doi:10.1007/978-0-387-77064-2_20

8. Magill RA, Anderson DI. *Motor Learning and Control: Concepts and Applications*. Vol 11. McGraw-Hill New York; 2007.

9. Tani G, Corrêa UC, Basso L, Benda RN, Ugrinowitsch H, Choshi K. An adaptive process model of motor learning: insights for the teaching of motor skills. *Nonlinear Dyn Psychol Life Sci*. 2014;18(1):47-65.

10. Schöner G, Zanone PG, Kelso JA. Learning as change of coordination dynamics: theory and experiment. *J Mot Behav*. 1992;24(1):29-48. doi:10.1080/00222895.1992.9941599

11. Reis J, Schambra HM, Cohen LG, et al. Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proc Natl Acad Sci*. 2009;106(5):1590-1595. doi:10.1073/pnas.0805413106

12. Jacobs DM, Michaels CF. Direct learning. *Ecol Psychol*. 2007;19(4):321-349. doi:10.1080/10407410701432337

13. de Pinho ST, Letícia da Silva S, Clavijo FAR, Alves D, Corrêa UC. The learning of decision-making on interception and passing in futsal. *Int J Sport Exerc Psychol*. 2021;19(6):1005-1021. doi:10.1080/1612197X.2020.1803950

14. Ranganathan R, Lee MH, Newell KM. Repetition without repetition: Challenges in understanding behavioral flexibility in motor skill. *Front Psychol*. 2020;11:2018. doi:10.3389/fpsyg.2020.02018

15. Zanone PG, Kelso JA. Evolution of behavioral attractors with learning: nonequilibrium phase transitions. *J Exp Psychol Hum Percept Perform*. 1992;18(2):403-421.

16. Adams JA. A closed-loop theory of motor learning. *J Mot Behav*. 1971;3(2):111-150. doi:10.1080/00222895.1971.10734898

17. Russell DM, Newell KM. On No-KR tests in motor learning, retention and transfer. *Hum Mov Sci*. 2007;26(1):155-173. doi:10.1016/j.humov.2006.07.009

18. Swinnen SP. Information feedback for motor skill learning: A review. In: Zelaznik HN, ed. *Advances in Motor Learning and Control*. Human Kinetics; 1996:37-66.

19. Levac DE, Huber ME, Sternad D. Learning and transfer of complex motor skills in virtual reality: a perspective review. *J Neuroengineering Rehabil*. 2019;16(1):121. doi:10.1186/s12984-019-0587-8

20. Haar S, van Assel CM, Faisal AA. Motor learning in real-world pool billiards. *Sci Rep*. 2020;10(1):20046. doi:10.1038/s41598-020-76805-9

21. Müller H, Sternad D. Decomposition of variability in the execution of goal-oriented tasks: three components of skill improvement. *J Exp Psychol Hum Percept Perform*. 2004;30(1):212-233. doi:10.1037/0096-1523.30.1.212

22. Sternad D, Abe MO, Hu X, Müller H. Neuromotor noise, error tolerance and velocity-dependent costs in skilled performance. *PLoS Comput Biol*. 2011;7(9):e1002159. doi:10.1371/journal.pcbi.1002159

23. McDonald PV, Oliver SK, Newell KM. Perceptual-motor exploration as a function of biomechanical and task constraints. *Acta Psychol (Amst)*. 1995;88(2):127-165. doi:10.1016/0001-6918(93)E0056-8

24. Pacheco MM, Lafe CW, Newell KM. Search strategies in the perceptual-motor workspace and the acquisition of coordination, control, and skill. *Front Psychol*. 2019;10:1874. doi:10.3389/fpsyg.2019.01874

25. Wulf G. Attentional focus and motor learning: a review of 15 years. *Int Rev Sport Exerc Psychol*. 2013;6(1):77-104. doi:10.1080/1750984X.2012.723728

26. Carlton LG. The effects of temporal-precision and time-minimization constraints on the spatial and temporal accuracy of aimed hand movements. *J Mot Behav*. 1994;26(1):43-50. doi:10.1080/00222895.1994.9941660

27. Zelaznik HN, Mone S, McCabe GP, Thaman C. Role of temporal and spatial precision in determining the nature of the speed-accuracy trade-off in aimed-hand movements. *J Exp Psychol Hum Percept Perform*. 1988;14(2):221-230. doi:10.1037/0096-1523.14.2.221

28. Lokesh R, Ranganathan R. Haptic Assistance that restricts the use of redundant solutions is detrimental to motor learning. *IEEE Trans Neural Syst Rehabil Eng*. 2020;28(6):1373-1380. doi:10.1109/TNSRE.2020.2990129

29. Hsieh TY, Liu YT, Mayer-Kress G, Newell KM. The movement speed-accuracy relation in space-time. *Hum Mov Sci*. 2013;32(1):257-269. doi:10.1016/j.humov.2012.12.010

30. Fitts PM, Posner MI. *Human Performance*. Brooks/Cole; 1967.

31. Reynolds B. The effect of learning on the predictability of psychomotor performance. *J Exp Psychol*. 1952;44(3):189-198. doi:http://dx.doi.org/10.1037/h0060862

32. Welch M, Henry FM. Individual differences in various parameters of motor learning. *J Mot Behav*. 1971;3(1):78-96. doi:10.1080/00222895.1971.10734895

33. Ranganathan R, Scheidt RA. Organizing and reorganizing coordination patterns. In: Laczko J, Latash ML, eds. *Progress in Motor Control*. Advances in Experimental Medicine and Biology. Springer International Publishing; 2016:327-349. doi:10.1007/978-3-319-47313-0_18

34. Day KA, Leech KA, Roemmich RT, Bastian AJ. Accelerating locomotor savings in learning: compressing four training days to one. *J Neurophysiol*. 2018;119(6):2100-2113. doi:10.1152/jn.00903.2017 652

35. Baguley T. Standardized or simple effect size: what should be reported? *Br J Psychol Lond Engl*. 1953. 2009;100(Pt 3):603-617. doi:10.1348/000712608X377117

36. Caldwell A, Vigotsky AD. A case against default effect sizes in sport and exercise science. *PeerJ*. 2020;8:e10314. doi:10.7717/peerj.10314

37. Shea CH, Kohl RM. Specificity and variability of practice. *Res Q Exerc Sport*. 1990;61(2):169-177. doi:10.1080/02701367.1990.10608671

38. Guadagnoli MA, Lee TD. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *J Mot Behav*. 2004;36(2):212-224. doi:10.3200/JMBR.36.2.212-224 661

39. Bobrownicki R, Collins D, Sproule J, MacPherson AC. Redressing the balance: Commentary on "examining motor learning in older adults using analogy instruction" by tse, wong, and masters (2017). *Psychol Sport Exerc*. 2018;38:211-214. doi:10.1016/j.psychsport.2018.05.014

40. Bourne LE, Archer EJ. Time continuously on target as a function of distribution of practice. *J Exp Psychol*. 1956;51(1):25-33.

41. Schmidt RA, Lange C, Young DE. Optimizing summary knowledge of results for skill learning. *Hum Mov Sci*. 1990;9(3-5):325-348.

42. Bobrownicki R, Carson HJ, MacPherson AC, Collins D. Unloading the dice: selection and design of comparison and control groups in controlled trials to enhance translational impact within motor learning and control research. *Int J Sport Exerc Psychol*. 2021. doi:10.1080/1612197X.2021.1956567

43. Henry FM. Specificity vs. generality in learning motor skill. In: Brown RC, Kenyon GS, eds. *Classical Studies on Physical Activity*. Prentice Hall; 1968:331-340.

44. Krishnan C, Dharia AK, Augenstein TE, et al. Learning new gait patterns is enhanced by specificity of training rather than progression of task difficulty. *J Biomech.* 2019;88:33-37. doi:10.1016/j.jbiomech.2019.03.014

45. Wymbs NF, Bastian AJ, Celnik PA. Motor skills are strengthened through reconsolidation. *Curr Biol*. 2016;26(3):338-343. doi:10.1016/j.cub.2015.11.066

46. Cohen J. Things I have learned (so far). *Am Psychol*. 1990;45(12):1304-1312. doi:10.1037/0003-066X.45.12.1304

47. Lohse K, Buchanan T, Miller M. Underpowered and overworked: Problems with data analysis in motor learning studies. *J Mot Learn Dev*. 2016;4(1):37-58. doi:10.1123/jmld.2015-0010

48. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14(5):365-376. doi:10.1038/nrn3475

49. Makin TR, Orban de Xivry JJ. Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. Rodgers P, Parsons N, Holmes N, eds. *eLife*. 2019;8:e48175. doi:10.7554/eLife.48175

50. Lakens D. Sample size justification. *Collabra:Psychology*. 2022; 8(1):33267. doi:10.1525/collabra.33267.

51. McKay B, Hussien J, Carter M, Yantha Z, Ste-Marie D. Expecting to teach a novel golf putting task did not enhance retention performance: A replication attempt of Daou and colleagues. *Commun Kinesiol*. 2021;1(2). doi:10.51224/cik.v1i2.39

52. Kantak SS, Winstein CJ. Learning-performance distinction and memory processes for motor skills: a focused review and perspective. *Behav Brain Res*. 2012;228(1):219-231. doi:10.1016/j.bbr.2011.11.028

53. Cardis M, Casadio M, Ranganathan R. High variability impairs motor learning regardless of whether it affects task performance. *J Neurophysiol*. 2017;119(1):39-48. doi:10.1152/jn.00158.2017

54. Grand KF, Bruzi AT, Dyke FB, et al. Why self-controlled feedback enhances motor learning: Answers from electroencephalography and indices of motivation. *Hum Mov Sci*. 2015;43:23-32. doi:10.1016/j.humov.2015.06.013

55. Hauser DJ, Ellsworth PC, Gonzalez R. Are manipulation checks necessary? *Front Psychol*. 2018;9:998. doi:10.3389/fpsyg.2018.00998

56. Taubert M, Draganski B, Anwander A, et al. Dynamic properties of human brain structure: Learning-related changes in cortical areas and associated fiber connections. *J Neurosci*. 2010;30(35):11670-11677. doi:10.1523/JNEUROSCI.2567-10.2010

57. Adams JA, Reynolds B. Effect of shift in distribution of practice conditions following interpolated rest. *J Exp Psychol*. 1954;47(1):32-36. doi:10.1037/h0061881

58. Adams JA. Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychol Bull*. 1987;101(1):41-74. doi:10.1037/0033-2909.101.1.41

59. Fischman MG, Christina RW, Vercruyssen MJ. Retention and transfer of motor skills: A review for the practitioner. *Quest*. 1981;33(2):181-194. doi:10.1080/00336297.1981.10483753

60. Van Rossum JHA. Schmidt's schema theory: the empirical base of the variability of practice hypothesis. *Hum Mov Sci*. 1990;9:387-435.

61. Wulf G, Schmidt RA. Variability in practice. *J Mot Behav*. 1988;20(2):133-149. doi:10.1080/00222895.1988.10735438

62. Hall KG, Magill RA. Variability of practice and contextual interference in motor skill learning. *J Mot Behav*. 1995;27(4):299-309. doi:10.1080/00222895.1995.9941719

63. Simmons JP, Nelson LD, Simonsohn U. False-Positive Psychology: Undisclosed flexibility in data Collection and analysis allows presenting anything as significant. *Psychol Sci*. 2011;22(11):1359-1366. doi:10.1177/0956797611417632

64. Fischman MG. On the continuing problem of inappropriate learning measures: Comment on Wulf et al. (2014) and Wulf et al. (2015). *Hum Mov Sci*. 2015;42:225-231. doi:10.1016/j.humov.2015.05.011

65. Hancock GR, Butler MS, Fischman MG. On the problem of two-dimensional error scores: Measures and analyses of accuracy, bias, and consistency. *J Mot Behav*. 1995;27(3):241-250. doi:10.1080/00222895.1995.9941714

66. Newell KM. More on absolute error, etc. *J Mot Behav*. 1976;8(2):139-142. doi:10.1080/00222895.1976.10735064

67. Krishnan C, Ranganathan R, Kantak SS, Dhaher YY, Rymer WZ. Active robotic training improves locomotor function in a stroke survivor. *J NeuroEngineering Rehabil*. 2012;9(1):57. doi:10.1186/1743-0003-9-57

68. Wulf G, Raupach M, Pfeiffer F. Self-controlled observational practice enhances learning. *Res Q Exerc Sport*. 2005;76(1):107-111. doi:10.1080/02701367.2005.10599266

69. Schutz RW, Roy EA. Absolute error: The devil in disguise. *J Mot Behav*. 1973;5(3):141-153. doi:10.1080/00222895.1973.10734959

70. Lee MH, Farshchiansadegh A, Ranganathan R. Children show limited movement repertoire when learning a novel motor skill. *Dev Sci*. 2018;21(4):e12614. doi:10.1111/desc.12614

71. Lee TD, Genovese ED. Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Res Q Exerc Sport*. 1988;59(4):277-287. doi:10.1080/02701367.1988.10609373

72. Schmidt RA. The Case against learning and forgetting scores. *J Mot Behav*. 1972;4(2):79-88. doi:10.1080/00222895.1972.10734922

73. Netz Y, Lidor R, Ziv G. Small samples and increased variability – discussing the need for restricted types of randomization in exercise interventions in old age. *Eur Rev Aging Phys Act*. 2019;16(1):17. doi:10.1186/s11556-019-0224-3

74. Trowman R, Dumville JC, Torgerson DJ, Cranny G. The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study. *J Clin Epidemiol*. 2007;60(12):1229-1233. doi:10.1016/j.jclinepi.2007.03.014

75. Mandava P, Kalkonde YV, Rochat RH, Kent TA. A matching algorithm to address imbalances in study populations: application to the National Institute of Neurological Diseases and Stroke Recombinant Tissue Plasminogen Activator acute stroke trial. *Stroke*. 2010;41(4):765-770. doi:10.1161/STROKEAHA.109.574103

76. Fu R, Holmer HK. *Change Score or Followup Score? An Empirical Evaluation of the Impact of Choice of Mean Difference Estimates*. Agency for Healthcare Research and Quality (US); 2015. Accessed December 19, 2021. http://www.ncbi.nlm.nih.gov/books/NBK285983/

77. Senn S. Change from baseline and analysis of covariance revisited. *Stat Med*. 2006;25(24):4334-4344. doi:10.1002/sim.2682

78. Krishnan C, Washabaugh EP, Reid CE, Althoen MM, Ranganathan R. Learning new gait patterns: Age-related differences in skill acquisition and interlimb transfer. *Exp Gerontol*. 2018;111:45-52. doi:10.1016/j.exger.2018.07.001

79. Lee MH, Ranganathan R. Age-related deficits in motor learning are associated with altered motor exploration strategies. *Neuroscience*. 2019;412:40-47. doi:10.1016/j.neuroscience.2019.05.047

80. Winstein CJ, Merians AS, Sullivan KJ. Motor learning after unilateral brain damage. *Neuropsychologia*. 1999;37(8):975-987. doi:10.1016/S0028-3932(98)00145-6

81. Lord FM. A paradox in the interpretation of group comparisons. *Psychol Bull*. 1967;68(5):304-305. doi:10.1037/h0025105

82. Bönstrup M, Iturrate I, Thompson R, Cruciani G, Censor N, Cohen LG. A rapid form of offline consolidation in skill learning. *Curr Biol*. 2019;29(8):1346-1351.e4. doi:10.1016/j.cub.2019.02.049

83. Dayan E, Cohen LG. Neuroplasticity subserving motor skill learning. *Neuron*. 2011;72(3):443-454. doi:10.1016/j.neuron.2011.10.008

84. Karni A, Meyer G, Rey-Hipolito C, et al. The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proc Natl Acad Sci U S A*. 1998;95(3):861-868. doi:10.1073/pnas.95.3.861

85. Lee MH. Online and offline contributions to motor learning change with practice, but are similar across development. *Exp Brain Res*. 2019;237(11):2865-2873. doi:10.1007/s00221-019-05639-3

86. Buxton CE. The status of research in reminiscence. *Psychol Bull*. 1943;40(5):313-340. doi:10.1037/h0053614

87. Pan SC, Rickard TC. Sleep and motor learning: Is there room for consolidation? *Psychol Bull*. 2015;141(4):812-834. doi:10.1037/bul0000009

88. Heathcote A, Brown S, Mewhort DJ. The power law repealed: the case for an exponential law of practice. *Psychon Bull Rev*. 2000;7(2):185-207.

89. Newell KM, Liu YT, Mayer-Kress G. Time scales in motor learning and development. *Psychol Rev*. 2001;108(1):57-82.

90. Newell KM, Mayer-Kress G, Liu YT. Human learning: Power laws or multiple characteristic time scales? *Tutor Quant Methods Psychol*. 2006;2(2):66-76.

91. Hawe RL, Scott SH, Dukelow SP. Taking proportional out of stroke recovery. *Stroke*. 2018;50:204-211. doi:10.1161/STROKEAHA.118.023006

92. Senn S. Statistical pitfalls of personalized medicine. *Nature*. 2018;563(7733):619-621. doi:10.1038/d41586-018-07535-2

93. Robertson EM. Skill Memory: Mind the ever-decreasing gap for offline processing. *Curr Biol*. 2019;29(8):R287-R289. doi:10.1016/j.cub.2019.03.007

94. Smeets JBJ, Louw S. The contribution of covariation to skill improvement is an ambiguous measure: Comment on Müller and Sternad (2004). *J Exp Psychol Hum Percept Perform*. 2007;33(1):246-249. doi:10.1037/0096-1523.33.1.246

95. Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond bar and line graphs: Time for a new data presentation paradigm. *PLOS Biol*. 2015;13(4):e1002128. doi:10.1371/journal.pbio.1002128

96. Munafò MR, Nosek BA, Bishop DVM, et al. A manifesto for reproducible science. *Nat Hum Behav*. 2017;1(1):1-9. doi:10.1038/s41562-016-0021

97. Winstein CJ, Lewthwaite R, Blanton SR, Wolf LB, Wishart L. Infusing motor learning research into neurorehabilitation practice: a historical perspective with case exemplar from the accelerated skill acquisition program. *J Neurol Phys Ther JNPT*. 2014;38(3):190-200. doi:10.1097/NPT.0000000000000046

98. Ranganathan R, Doherty C, Gussert M, Kaplinski E, Koje M, Krishnan C. Scientific basis and active ingredients of current therapeutic interventions for stroke rehabilitation. *Restorative Neurology and Neuroscience*; 2022; 40(2):97-107. doi: 10.3233/RNN-211243