Observing different model types interspersed with physical practice has no effect on consolidation or motor learning of an elbow flexion–extension task

Clara M. Moore, Natasha Lelievre, Diane M. Ste-Marie*

School of Human Kinetics, University of Ottawa, Ottawa, ON K1N 1A2, Canada

ABSTRACT

We compared varied model types and their potential differential effects on learning outcomes and consolidation processes when observational practice was interspersed with physical practice. Participants (N = 75) were randomly assigned to one of five groups: (1) unskilled model observation, (2) skilled model observation, (3) mixed-model observation, (4) physical practice only, and (5) no observational or physical practice (control). All were tasked with learning a waveform-matching task. With exception of the control group not involved in acquisition sessions, participants were involved in one pre-test, two acquisition sessions, four retention tests (immediate-post acquisition 1, 24hr post acquisition 1, immediate-post acquisition 2, and approximate 7-day retention), as well as an approximate 7-day transfer test. No differences were demonstrated in consolidation processes or learning outcomes as all groups showed the same pattern of retention and transfer data. Our conclusion is that motor memory processes were not impacted differentially when different models types were used in observational practice that was intermixed with physical practice for the learning of a movement pattern with low task difficulty, and thus similar learning outcomes emerged for all groups.

1. Introduction

Motor skill learning has been described as occurring through a three-step motor memory process, which involves the processes that occur during practice of the skill (encoding), post-practice processes (consolidation), and those processes involved in recovering the necessary information to execute the motor skill after a time of no practice (retrieval; Robertson, 2009). Within the context of motor skill acquisition, there are several practice methods that can be used in order to learn a novel motor skill, and a number of these have been shown to enhance the efficacy of practice. Variations in practice scheduling (Lee & Magill, 1983; Shea & Morgan, 1979), frequency of feedback (Winstein & Schmidt, 1990; Wulf, Schmidt, & Deubel, 1993), and the provision of action observation (Heyes & Foster, 2002; Lee & White, 1990; Starek & McCullagh, 1999) are examples of some of these practice variables. Of interest to this research is the use of action observation (also referred to as observational learning), which allows learners to acquire the ability to perform an action by observing themselves or others executing the task to be learned (McCullagh, Ste-Marie, & Law, 2013). More specifically, our interest lies in the effects of observational practice, in conjunction with physical practice, on motor memory consolidation and subsequent learning outcomes.

Consolidation, the second stage of motor memory formation, is said to only begin once the learner has stopped physically

---

practicing the skill and it involves post-practice processes that help to transfer information into the long-term over the passage of time (Kantak & Winstein, 2012; Trempe, Sabourin, Rohbanfard, & Proteau, 2011). Consolidation has been found to have two distinct behavioral outcomes: performance stabilization and offline learning (Robertson, Pascual-leone, & Miall, 2004). Performance stabilization is characterized as a motor memory representation that becomes less susceptible to interference when other motor tasks are acquired shortly after acquisition of a preceding skill (e.g., Krakauer & Shadmehr, 2006) and/or, shows retention performance that is similar to the motor performance that was seen at the end of acquisition (Kantak, Sullivan, Fisher, Knowlton, & Winstein, 2010). Offline learning is different from performance stabilization because a spontaneous improvement in motor performance is observed after a time delay, often via performance testing after a 24-h period, without any further physical practice having occurred (Kantak & Winstein, 2012). Our specific interest was in performance stabilization, as detected by performance on 24-h or 7-day delayed retention tests, as well as offline learning processes that can be observed when comparing immediate retention performance to that of 24-h or 7-day delayed retention testing.

Although many experiments have investigated these two behavioral outcomes of consolidation following physical practice of a motor skill (Albouy et al., 2013; Brashers-Krug, Shadmehr, & Bizzi, 1996; Malangré, Leinen, & Blischke, 2014), there is relatively little research that has considered the effects of observational practice and its effects on consolidation. To our knowledge, the first experiments to examine observational learning and consolidation have shown varied results. For example, Van Der Werf, Van Der Helm, Schoonheim, Ridderikhoff, and Van Someren (2009) completed four experiments investigating consolidation outcomes following the observation of a finger-tapping task in conjunction with how the timing of sleep affected consolidation. The first experiment allowed participants to move their fingers in order to receive motor feedback while they were observing the task; however, they were not allowed to practice the actual sequence. In the following three experiments, the finger movement during observation was disallowed and manipulations were performed to investigate whether there were differences in consolidation when the sequence was observed in the evening compared to the morning. The overall finding from Van der Werf et al.'s research showed that there were offline learning benefits in adults learning a finger-tapping sequence, but only if observation was followed by sleep.

van Schalkwijk et al. (2015) implementing a similar experiment to Van Der Werf et al. (2009), had children learn a finger-tapping task through observation and no overt movements of the fingers were allowed. These researchers also investigated whether immediate sleep was required in order for consolidation outcomes to occur. Similar to Van Der Werf et al. (2009), it was found that sleep following observation was an important factor in children learning a finger-tapping sequence. Offline benefits, however, were not observed and instead benefits were seen in terms of performance stabilization, as measured by a delayed retention test. Trempe and colleagues (2011) also investigated the effects of observation on offline learning and performance stabilization in adults attempting to learn a knock-down barrier task. Again, it was found that observation alone led to performance stabilization, but not to offline learning. Despite the varied findings, these research findings do provide evidence that observational practice alone can impact consolidation processes and that sleep is of importance in any experimental design investigating consolidation effects.

None of the research to date, to our knowledge, however, has considered the effects of observational and physical practice interspersed within a motor skill acquisition session. This is of importance, because it has been shown that practice structures which combine both observational and physical practice are more efficacious than a practice structure which isolates the two forms of practice (e.g., Shea, Wright, Wulf, & Whitacre, 2000). Indeed, Trempe et al. (2011) stated that further work was necessary to determine the combined effects of physical practice and observational practice with consolidation processes. Thus, understanding the impact of an interspersed observational/physical practice structure on consolidation warrants investigation and heeds Trempe et al. (2011) recommendations.

Of further relevance is that the experimental methods used to date have incorporated different models to demonstrate the to-be-learned motor task. Trempe et al. (2011) used an expert model, whereas Van Der Werf et al. (2009) and van Schalkwijk et al. (2015) both used learning models (i.e., a video that showed the first six attempts of the finger-tapping sequence by a person who was naïve to the task). We argue that consideration needs to be given to the model types used because different model types have been contended to generate learning benefits via different mechanisms (e.g., Rohbanfard & Proteau, 2011; Robertson, St. Germain, & Ste-Marie, 2018). As examples, skilled models are those that display the correct behaviours with minimal errors and are thought to be beneficial because they provide a reference of correctness of the movement patterning (Al-Abood, Davids, Bennett, Ashford, & Marin, 2001; Heyes & Foster, 2002). Learning models, however, display the skill in the early stages of learning, and thus, with more error in the motor performance, and progress to a higher skill level. Learning models, along with unskilled models (which only show errorful performance) are described as being beneficial to motor learning because they engage problem-solving processes that assists the learners with developing error detection and correction mechanisms associated with the movement (Buchanan & Dean, 2010; Lee & White, 1990). Further, because skilled and learning/unskilled models have been argued to be advantageous for different reasons, it has been proposed that combining these two different model types, termed mixed-modeling, could provide additive learning benefits. Research has shown benefits for mixed-models as compared to the use of just one model type and researchers have argued that this might be because the mixed-model observation afforded learners with the benefits specific to each model type (Rohbanfard & Proteau, 2011; Robertson et al., 2018).

In sum, considering there is little research regarding the combined effects of observational and physical practice on consolidation, and that the effect of model type has yet to be considered, our research goals were to replicate the mixed-modeling advantage and to determine if this learning advantage could be explained by differential consolidation processes. To this end, we adopted a similar experimental paradigm as that of Rohbanfard and Proteau (2011) and compared five experimental groups comprised of participants who (1) observed an unskilled model only (only poor performances of the skill shown) interspersed with physical practice, (2) observed a skilled model interspersed with physical practice, (3) observed a mixed-model that combined both the unskilled and skilled model types interspersed with physical practice, (4) performed physical practice only, and (5) simply read a magazine (control

group with no physical or observational practice).

We also explored the impact of two acquisition sessions on different days as most research has only included one acquisition session, and it can easily be argued that most skill acquisition can occur across multiple days of practice. As little is known regarding multiple acquisition sessions, we were uncertain how consolidation and motor learning would be affected following the second day of acquisition, and thus, no specific hypotheses are forwarded. We did hypothesize, however, that those in the mixed-model observation group would show the superior learning outcomes as compared to the other modeling groups and physical practice only. Although we anticipated these superior learning outcomes for the mixed-modeling group over the other modeling groups and the control, we were uncertain as to whether such differences would be driven by motor memory consolidation processes. An examination of performance changes between immediate and delayed retention tests was expected to provide us with the opportunity to explore the possible contributions of consolidation processes.

2. Method

2.1. Participants

Data were collected from 77 university-aged individuals (M_{age} = 21.6, SD = 1.97, M/F: 26/51). Of these, two participants were removed from the analyses due to missing data, these two participants were replaced with two new participants who were put into the missing participants' original experimental groups resulting in 75 participants completing the entire protocol. Using self-report, it was determined that participants had no prior experience with the task and had no sensory or motor dysfunctions. All participants were determined to be right-handed with the Edinburgh Handedness Inventory (Oldfield, 1971) and provided written informed consent. The experimental protocol received approval from the University's Research Ethics Board.

2.2. Task and apparatus

Participants were required to learn a waveform-matching task, wherein they used two elbow extension and flection movements with their non-dominant (left) arm to replicate a waveform as accurately as possible with the goal movement time of 900 ms. The waveform was created by summing two sine waves: $y(t) = 42\sin(\pi t - 0.3) + 23\sin(3\pi t + 0.4)$. Participants sat in a chair facing a computer monitor displaying the waveform task and their left forearm rested on a padded armrest attached to a manipulandum that they grasped with their left hand (Fig. 1). The manipulandum was also attached to an axis restricting movement to the horizontal plane (the movements were restricted to the elbow joint). A linear potentiometer, powered by a 5 V direct current power supply, was attached to the central axis of the manipulandum to provide position data relating to the participant's arm movements. A customized LabVIEW (National Instruments Inc.) program was used to control the timing of all experimental stimuli and recorded and stored the data for analysis.

2.3. Materials

Three different modeling videos were used throughout the experiment, specifically, (1) a peer-skilled model video, (2) a peer-unskilled model video, and (3) a mixed-model video (combination of peer-unskilled and peer-skilled model). The peer-skilled and
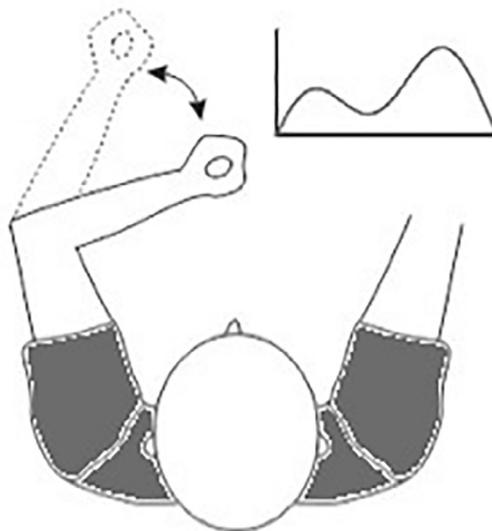

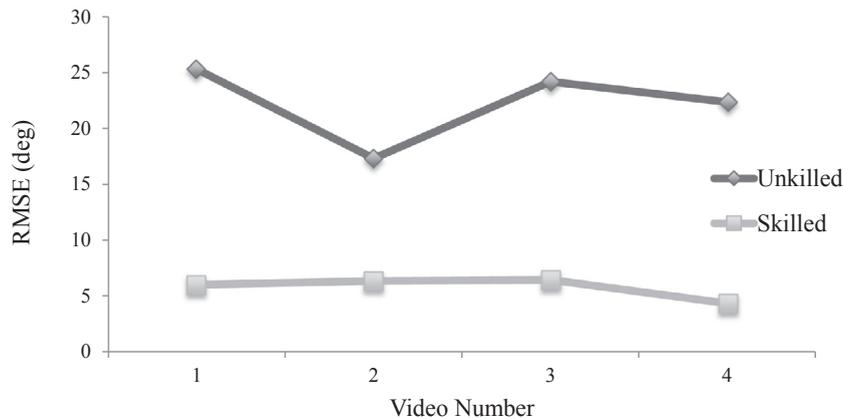**Fig. 1.** A visual representation of the required movement and the target waveform.

Fig. 2. The RMSE of the model in the videos that were watched by the participants. Those in the unskilled group and skilled group observed video numbers 1 and 2 during the first observation session in a block and observed videos 3 and 4 in their second observation session in a block. Participants in the mixed-model group observed unskilled video 1 and skilled video 1 during their first observation session and unskilled video 2 and skilled video 2 during their second.

peer-unskilled model videos each showed four trials of the respective model performing the task. The mixed-model video also showed four trials but alternated between a peer-skilled model trial and a peer-unskilled model trial. All video conditions were filmed in the subjective viewpoint, where the camera was placed behind the model performing the task (Ste-Marie et al., 2012). The modeling videos not only showed the model's arm performing the movement but also showed the feedback the model received via the computer screen once the movement was complete.

The skilled model and unskilled model was the same person, who was a peer (i.e., male student of same approximate age as participants). This person was videoed for the first 10 trials of practice with the task. These 10 trials were then analyzed to determine the trials that had the worst root mean square error (RMSE) and the worst movement times. The four worst trials were then selected to create the unskilled model videos provided during the acquisition phase of the experiment (refer to Figs. 2 and 3). The person then continued practicing over the time course of several days (performing approximately 150 trials total) and when he was able to execute the task with greater consistency at a near perfect performance, he was video-recorded for 40 trials. From those 40 trials, the four best trials were selected in terms of RMSE and the movement time performance to create the video sequences of the skilled model component (refer to Figs. 2 and 3).

2.4. Procedure

Participants were randomly assigned to one of five experimental groups (n = 15 per group): Mixed-model observation (MM), Unskilled model observation (UM), Skilled model observation (SM), Physical practice only (PP) or a control (C). All participants completed three days of testing and were informed that they would be asked to report their hours of sleep across the days of the week at the end of the experimental phases. Participants completed their testing sessions any time between 8 am and 8 pm.
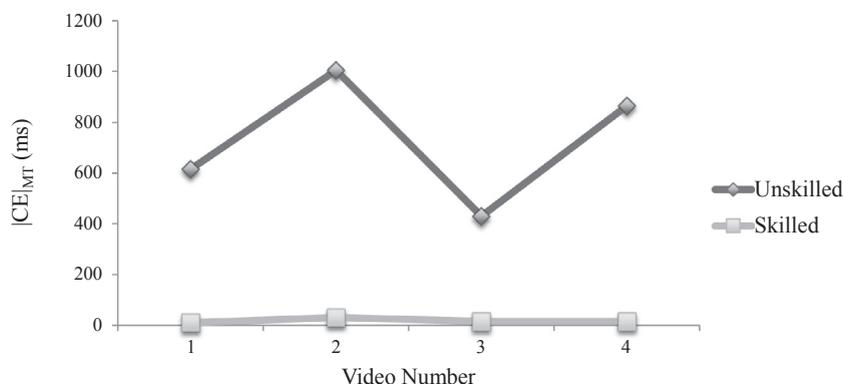

Fig. 3. The absolute constant movement time error of the model in the videos that were watched by the participants. Those in the unskilled group and skilled groups observed videos numbers 1 and 2 during the first observation session in a block and observed videos 3 and 4 in their second observation session in a block. Participants in the mixed-model group observed unskilled video 1 and skilled video 1 during their first observation session and unskilled video 2 and skilled video 2 during their second.
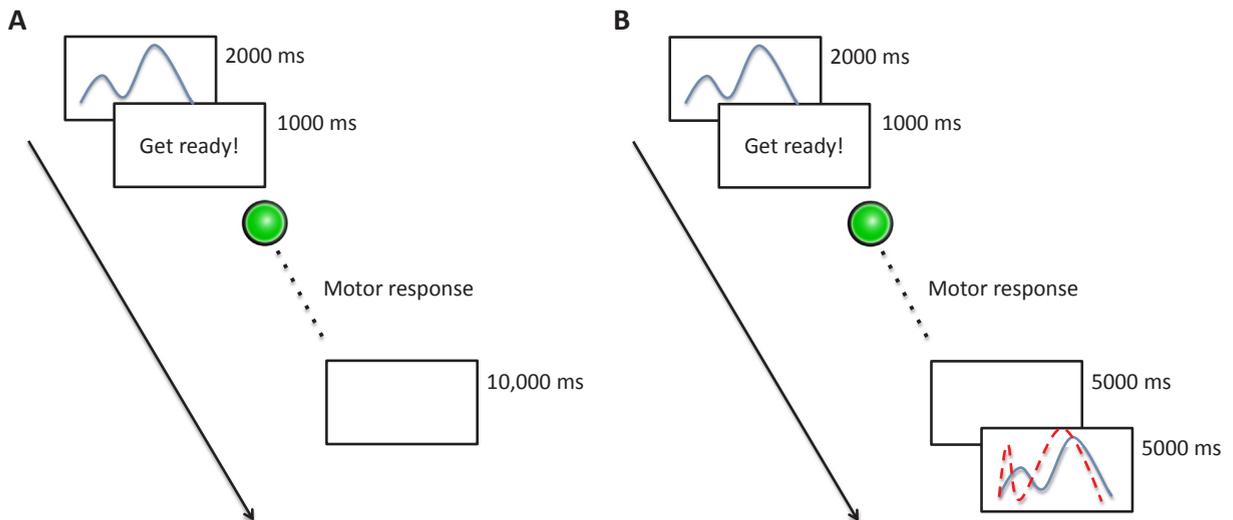
Fig. 4. Schematic of the timing of events during the pre-test, retention and transfer tests (A) and during a typical acquisition trial (B).

2.4.1. Day 1

Participants in all groups began Day 1 with a pre-test consisting of 10 physical practice trials of the skill where they received no feedback regarding their performance. Prior to beginning the trials, participants were informed that when they moved their arm away from their body, it would make the line of the waveform go up, and when the arm was moved back towards the body, the wave form line would go down. Each trial began with the image of the target waveform displayed on the screen for two seconds, followed by a visual "get ready" and a visual "go" cue, in which a light turned green (Fig. 4). As there was no interest in reaction time the participants were able to begin their movement whenever they liked following the "go" cue. Once the participant began their movement the computer screen remained blank. Once their movement was complete there was a 10 s delay before the next trial.

Subsequent to the pre-test, those in the observation groups (MM, UM, SM) completed the first acquisition session, which consisted of 9 blocks of 6 physical practice trials and 4 observational practice trials of the specific modeling video. Within each block, physical and observational practice were interspersed in that participants performed three physical practice trials followed by observing two modeling videos, this was then repeated a second time for a total of six physical and four observational practice trials. Those in the physical practice only group also had 9 blocks of 10 trials, but no observational practice was provided; all trials involved physical execution of the task.

During acquisition sessions, all participants received augmented feedback regarding their physical performance after a 5 s-delay. The augmented feedback was displayed for 5 s and was comprised of the waveform executed by the participant superimposed on the correct waveform. The movement time was also presented in the top right-hand corner of the screen (see Fig. 4 for sequencing of trials). Following Session 1 of acquisition, all participants in the PP, UM, SM and MM groups took a 10-minute break. Participants in the control group did not complete Acquisition 1, instead they read magazines for 40 min, which resulted in the same delay-interval as those in the other groups. Finally, all participants completed a retention test (Immediate-post acquisition 1 retention; I-PA1R) in which they performed 10 physical practice trials with no feedback.

2.4.2. Day 2

Participants were asked to come to Day 2 of data acquisition that took place 24-h following the initial acquisition session. Participants in all groups began with a 24 h delayed retention test (24-PA1R), which consisted of 10 trials with no feedback provided. Then all participants, excluding those in the control group, followed the exact same procedure as they did on Day 1; that is, 90 acquisition trials were provided, followed by a 10 min immediate retention test (immediate-post acquisition 2 retention; I-PA2R) for the modeling and physical practice only groups. The control participants only completed the two retention tests and again read magazines for the 40 min between the two tests.

2.4.3. Day 3

The final testing day was completed 6–8 days following Day 1. Day 3 consisted of one retention test (hereafter Weeklong Retention; WR), which followed the same procedures of the other retention tests, and a transfer test. The waveform used during the transfer test was the reverse of the waveform learned by participants during acquisition and had a new associated goal movement time of 1150 ms. Participants completed 10 trials of this task and no feedback was provided. Additionally, on the last day of testing, all participants were requested to complete a sleep diary indicating how many hours they slept each night during the past week.

2.5. Data

2.5.1. Dependent variables

Three dependent variables were used; (1) Root mean square error (RMSE), (2) temporal accuracy for the overall movement time; and spatial accuracy specific to the three reversal points of the movement. RMSE was calculated as the mean difference between the target waveform and the participant's response over the participant's actual movement time. RMSE represents the participants' overall performance accuracy per trial and therefore includes spatial and temporal components (Carter & Ste-Marie, 2016). RMSE has been shown to be advantageous for two reasons: not only is it sensitive to spatial and temporal errors in the participant's motor response compared to the target waveform, but also, RMSE incorporates variability and bias of the performed response (Kovacs, Boyle, Grutmatcher, & Shea, 2010; Schmidt & Lee, 2011). The temporal accuracy of the overall movement time was determined by using the absolute constant error (|CE|) of the movement time with respect to the goal movement time (Carter & Ste-Marie, 2017; Lin et al., 2009), see (Fig. 5). The spatial accuracy was determined by using the sum of the |CE| in movement amplitude for each reversal point on the waveform (Carter et al., 2018; Lin et al., 2009), see (Fig. 5).

2.5.2. Data analysis

Descriptive statistics were run to explore all the data. Two participants were not included in the analyses as they had missing data due to the customized LabVIEW program not saving the files properly; these participants were replaced. Therefore, the data for 75 participants were used for all analyses. We assessed the normality of the distribution by assessing skewness and kurtosis of the standardized residuals as well as performing the Kolmogorov-Smirnov test. We also performed Levene's test to assess homogeneity of variance. There were several instances where the assumptions of normality of homogeneity of variance were violated. Norman (2010), however, argues that, for an analysis of variance (ANOVA), the assumption of normality speaks to the normality of the distribution of means, not the distribution of the data. Indeed, the Central Limit Theorem demonstrates that for sample sizes larger than 5 or 10 in a group, the means will be normally distributed, no matter the original distribution (Norman, 2010). Thus, although the data was not normally distributed in all cases there were no transformations[1], winsorizing or trimming performed. Box plots were used to detect any possible outliers and three outliers were removed from all analyses, as they were extreme cases in terms of every dependent variable. Extreme cases were defined as values greater than 3 times the inter quartile range. This resulted in a final sample size of 72.

All data are expressed as means with standard deviations, and F values are provided for the main effects and interactions. Partial eta squared (η_p^2) is reported in order to give an estimate of the amount of variance that can be attributed to any significant main effects or interactions. If Mauchly's Test of Sphericity was violated a Greenhouse-Geisser correction or Huynh-Feldt correction was performed, depending whether epsilon was less or greater than 0.75. Statistical significance was set at $p < .05$. Bonferroni post-hoc testing was used to determine significant differences amongst means when main effects or interactions were found.

3. Results

3.1. Preliminary analysis

In order to ensure that participants in all groups began at the same skill level, a one-way ANOVA was conducted on the pre-test data for all three dependent variables. There were no significant differences found between groups for any of the dependent variables; RMSE, $F(4, 67) = 0.686$, $p = .604$ (Fig. 6), temporal accuracy ($p = .72$) (Fig. 7), and spatial accuracy ($p = .691$) (Fig. 8).

Two separate 5 Group (UM, SM, MM, PP, C) one-way ANOVAs were performed to determine if there were any significant differences in the amount of hours of sleep per night between groups prior to the 24-PA1R and WR/T tests. No significant group differences were found ($p > .05$) The average sleep data for all groups are presented in Table 1.

3.2. Acquisition

To examine the acquisition data, separate 4 Group (UM, SM, MM, PP) × 2 Session (Acquisition 1, Acquisition 2) × 9 Block (blocks 1–9) three-way, mixed ANOVAs with repeated measures on the last two factors were performed for each dependent variable.

3.2.1. Root mean square error

While there was a significant main effect for session, $F(1, 56) = 29.09$, $p < .000$, $\eta_p^2 = 0.342$ and block, $F(4.33, 242.625) = 33.879 p < .000$, $\eta_p^2 = 0.377$, both of these were superseded by a significant Session × Block interaction, $F(4.371, 244.754) = 16.185$, $p < .000$, $\eta_p^2 = 0.224$. Post-hoc testing indicated significant performance differences between blocks 1 and 2 in session 1 of acquisition ($p < .05$), continued improvement occurred as blocks 1 and 2 were significantly different from block 4 ($p < .05$) and then no performance changes occurred thereafter. In contrast, no performance differences occurred across blocks in session 2 ($p > .05$). There was no significant main effect for group ($\eta_p^2 = 0.045$). All RMSE results can be seen in Fig. 6.

---

[1] A log transformation did not change the significance of the findings, thus for ease of interpretation the non-transformed data was analyzed.
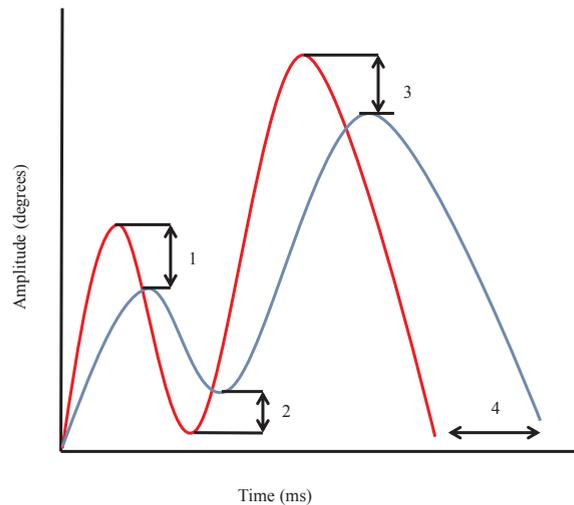
Fig. 5. Visual representation of the goal waveform compared with a participant's waveform. The blue line represents the target waveform and the red line represents the waveform drawn by a participant. Spatial accuracy was determined by summing the absolute constant error in movement amplitude, demonstrated by numbers 1 through 3 ($\Sigma|CE|_{Amp}$). Temporal accuracy was calculated by finding the absolute of the movement time difference between the target waveform and the waveform drawn, represented by number 4 ($|CE|_{MT}$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Mean number of hours slept before the 24 h post acquisition 1 retention test and the weeklong retention test with the standard deviations (SD) indicated in brackets.

Group	Day 1	Week
Unskilled model	6.47 (1.36)	7.14 (0.65)
Skilled model	7.17 (1.25)	7.43 (0.62)
Mixed-model	6.97 (1.17)	7.24 (0.71)
Physical practice	7.36 (1.78)	7.39 (0.99)
Control	7.1 (1.26)	7.16 (0.63)
Average	7.03 (1.36)	7.28 (0.72)

3.2.2. Temporal accuracy

The acquisition data again showed significant main effects for session, $F(1, 55) = 12.542$, $p = .001$, $\eta_p^2 = 0.186$, and block, $F(1.529, 84.071) = 21.015$, $p < .000$, $\eta_p^2 = 0.276$, yet these main effects were also superseded by a Session × Block interaction $F(1.779, 97.832) = 12.963$, $p < .000$, $\eta_p^2 = 0.191$. Post-hoc testing showed the same two features defining the interaction as that noted for RMSE. That is, performance changes only occurred within acquisition session 1, with significant differences seen between block 1 and all remaining blocks ($p < .05$), and the remaining blocks showing no performance changes ($p > .05$). No significant changes in performance, however, occurred across any of the blocks in session 2. As with RMSE there was no significant main effect for group ($\eta_p^2 = 0.031$). All temporal accuracy results can be seen in Fig. 7.

3.2.3. Spatial accuracy

Mirroring the RMSE and temporal accuracy, significant main effects for session and block, $F(1,55) = 20.744$, $p < .000$, $\eta_p^2 = 0.274$ and $F(2.443, 134.383) = 17.534$, $p < .000$, $\eta_p^2 = 0.242$, respectively, were found, but these were superseded by a significant Session × Block interaction, $F(3.032, 166.776) = 19.392$, $p < .000$, $\eta_p^2 = 0.261$. Post-hoc testing demonstrated a significant improvement from block 1 to block 2 in session 1 ($p < .05$), with no subsequent performance changes in the following blocks, whereas there were no demonstrated differences in performance across the blocks during session 2 ($p > .05$). Finally, there was no significant main effect for group ($\eta_p^2 = 0.087$). All spatial accuracy results can be seen in Fig. 8.

3.3. Retention

Retention data for all dependent variables were analyzed using three separate ANOVAS for each dependent variable. One omnibus, 5 Group (UM, SM, MM, PP, C) × 4 Time (IPA1R, 24-PA1R, IPA2R, WR) two-way mixed ANOVA with repeated measures on time. As well as a 5 Group (UM, SM, MM, PP, C) × 2 Time (I-PA1R, 24-PA1R) two way mixed ANOVA with repeated measures on time to investigate changes occurring over the 24-h retention period and a 5 Group (UM, SM, MM, PP, C) × 2 Time (I-PA2R, WR) with repeated measures on time to investigate changes over the weeklong retention period.
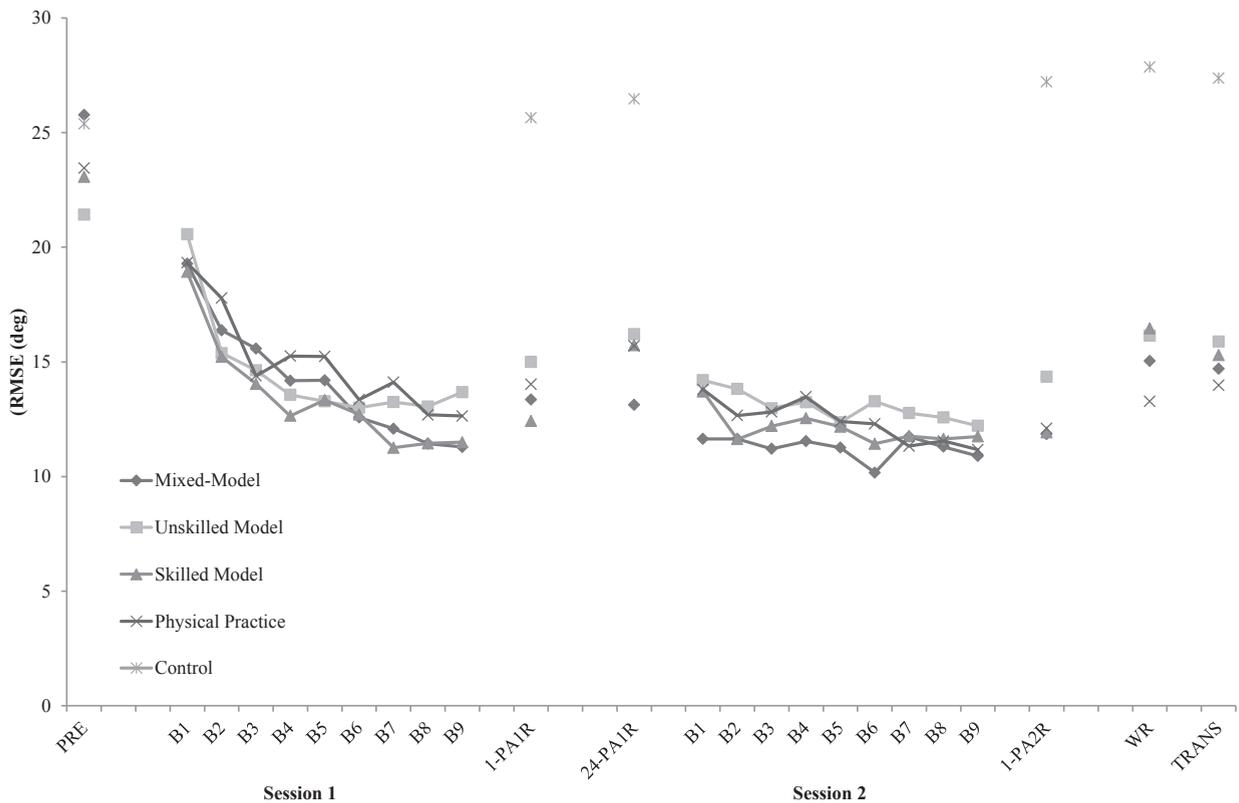
Fig. 6. Average RMSE scores in degrees for each group over the course of the experimental protocol.

3.3.1. Root mean square error

The omnibus ANOVA showed a significant main effect for time $F(2.896,194.052) = 7.878$, $p < .000$, $\eta_p^2 = 0.105$. Post-hoc testing demonstrated that the performance error for both immediate retention tests (I-PA1R/I-PA2R) were significantly lower than the weeklong retention. Further, the performance error scores during the 24-PA1R were significantly lower than the scores during the I-PA1R. No other means were significantly different. A main effect was also found for group $F(4, 67) = 36.388$, $p < .000$, $\eta_p^2 = 0.685$. Post-hoc testing showed that the control group was significantly worse in performance on all retention tests as compared to the four other groups, which were not different from each other.

Over the 24-h retention a significant main effect was found for time $F(1,67) = 7.109$, $p < .010$, $\eta_p^2 = 0.096$, thus demonstrating that the performance error for the 24-PA1R was significantly worse than the I-PA1R. A main effect was also found for group $F(4, 67) = 24.359$, $p < .000$, $\eta_p^2 = 0.593$. Post-hoc testing showed that the control group was significantly worse in performance on all retention tests as compared to the four other groups ($p < .000$), which were not different from each other ($p > .05$).

Similarly over the weeklong retention a significant main effect was found for time $F(1,67) = 19.356$, $p < .000$, $\eta_p^2 = 0.224$, demonstrating that the performance error for the WR was significantly worse than the I-PA2R. There was also a main effect for group $F(4, 67) = 32.651$, $p < .000$, $\eta_p^2 = 0.661$, where the post hoc showed that the control group was significantly worse than all other groups ($p < .000$).

3.3.2. Temporal accuracy

The omnibus ANOVA demonstrated significant main effects for both time and group, $F(1.428, 95.665) = 43.893$, $p = .037$, $\eta_p^2 = 0.055$ and $F(4, 67) = 12.332$, $p < .000$, $\eta_p^2 = 0.424$ respectively. It was found that I-PA1R was significantly worse than I-PA2R, however there were no significant differences between 1-PA1R and 24-PA1R or 1-PA2R and WR. Post-hoc testing for the main effect for group indicated that the control group was significantly worse than all other groups at all levels of retention, with no other differences emerging.

When looking at changes over the 24-h retention interval there was no significant main effect for Time ($p > .05$), however there was a significant Group by Time interaction, $F(4,67) = 3.206$, $p = .018$, $\eta_p^2 = 0.161$. The post-hoc showed that one of the modeling groups, specifically the SM group, was significantly better than both the PP and C at the I-PA1R ($p < .05$), whereas at the 24-h retention, all modeling groups were significantly better in performance that those in the PP and C ($p > .05$).

When investigating changes over the weeklong retention, the ANOVA demonstrated significant main effects for both time and group, $F(1,67) = 4.40$, $p = .040$, $\eta_p^2 = 0.062$ and $F(4,67) = 22.839$, $p < .000$, $\eta_p^2 = 0.577$ respectively. With respect to time, performance levels during the I-PA2R was significantly better than the WR ($p < .05$). In terms of the group main effect, post-hoc
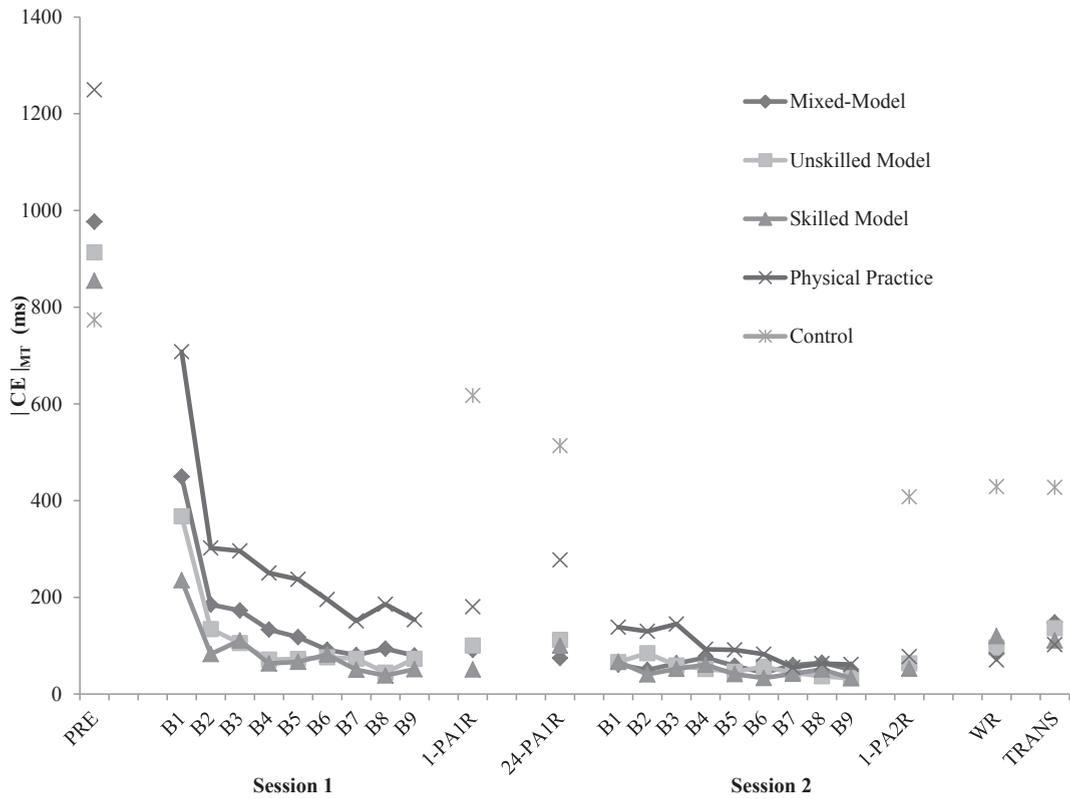
Fig. 7. Average temporal accuracy (absolute constant error) scores in milliseconds for each group over the course of the experimental protocol.
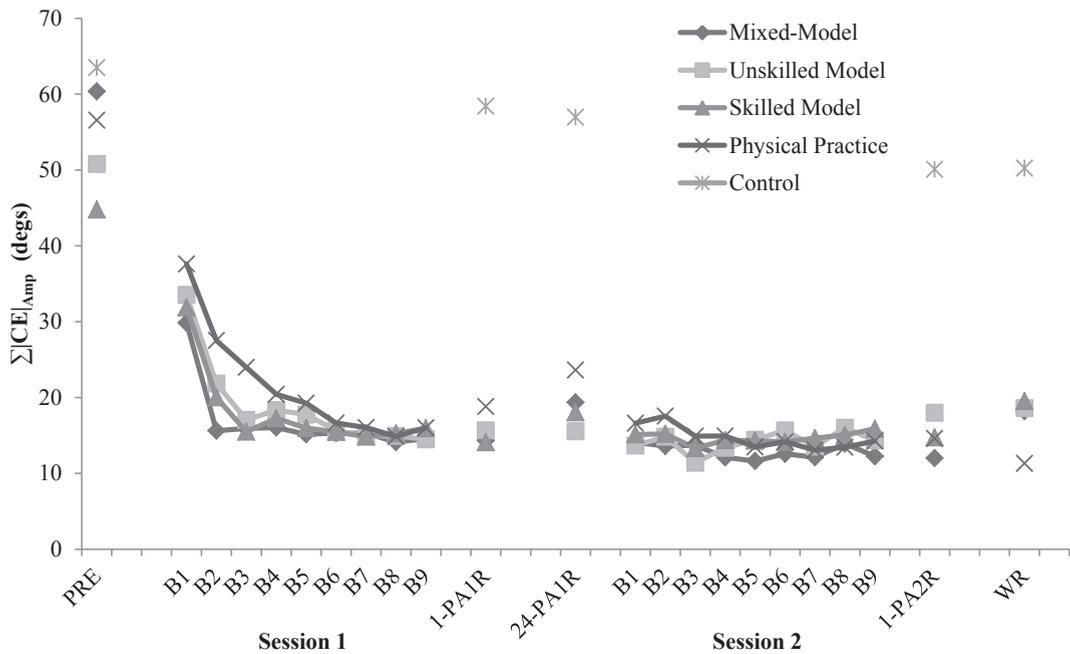

Fig. 8. Average spatial accuracy (absolute constant error) scores in milliseconds for each group over the course of the experimental protocol.

testing showed that the control group was significantly worse than all other groups ($p < .05$), with no other group differences ($p > .05$).

3.3.3. Spatial accuracy

In comparison to the two previous dependent variables, the omnibus ANOVA showed no main effect for time, indicating there were no significant differences among the obtained means across any of the retention sessions. The control group continued to perform significantly worse than all others, as evidenced by the post-hoc testing for the significant main effect for group $F(4, 67) = 17.253$, $p < .000$, $\eta_p^2 = 0.507$. There were no significant interactions as all the comparisons for mean spatial accuracy ($\Sigma|CE|_{Amp}$) were not statistically significant ($p > .05$).

As with temporal accuracy, over the 24-h retention, there was no main effect for time. The control group continued to perform significantly worse than all other groups, as evidenced by the post-hoc testing for the significant main effect for group $F(4,67) = 15.313$, $p < .000$, $\eta_p^2 = 0.478$. There were no significant interactions.

Over the weeklong retention, there was also no main effect for time and no significant interactions. As with all the other dependent variables, the control group performed significantly worse than all the other groups, as indicated by post-hoc testing; $F(4,67) = 14.428$, $p < .001$, $\eta_p^2 = 0.463$.

3.4. Transfer

A transfer test was included to investigate the adaptability of the newly acquired task, also giving insight into the amount of learning that occurred (Kantak & Winstein, 2012). To investigate the transfer data, separate 5 Group (UM, SM, MM, PP, C) one-way ANOVAs were performed for each of the dependent variables.

3.4.1. Root mean square error

There was a significant main effect for group, $F(4, 67) = 9.330$, $p < .000$, $\eta_p^2 = 0.358$ and post-hoc testing demonstrated that the control group was significantly worse than all other groups, none of which were different from each other.

3.4.2. Temporal accuracy

As with the RMSE, there was a significant main effect for group and the only difference, as evidenced by post-hoc testing, was that the control group was significantly worse than all other groups, $F(4, 67) = 8.768$, $p < .000$, $\eta_p^2 = 0.344$.

3.4.3. Spatial accuracy

There was a significant main effect for group, $F(4, 67) = 7.283$, $p < .000$, $\eta_p^2 = 0.303$ with post-hoc testing again showing that the control group was significantly worse than all other groups, which were not different from each other.

4. Discussion

While researchers have studied the use of observation and its effects on consolidation and learning outcomes (Trempe et al., 2011; Van Der Werf et al., 2009; van Schalkwijk et al., 2015), none have compared varied model types and their possible impact on learning, with a specific interest on consolidation processes. Consequently, the overall objectives of this experiment were to determine whether observation of a mixed-model in conjunction with physical practice, compared to a single model, would yield learning advantages that might be explained by differential consolidation processes. We also explored the possible impact of two acquisition sessions on these consolidation and learning process as the majority of research has only included one acquisition session.

Turning to the first objective, our findings showed no differences during acquisition or retention and transfer across the observation groups; thus, our results did not replicate the mixed-model advantage found by Rohbanfard and Proteau (2011). There were, however, some key differences between the two experimental designs that possibly explain these differing results. One example is that we used a true novice model (i.e., only errorful performances shown) versus their use of a learning model (performance improved across the observed trials). Likely of more importance, however, was that Rohbanfard and Proteau (2011) had their participants in the modeling groups only observe the task for 60 trials before the first immediate-retention test and then it was in a subsequent phase that those observation groups performed 60 physical practice trials. Our participants, however, had the observational and physical practice interspersed as soon as practice on the task began. Indeed, Rohbanfard and Proteau only found a learning advantage for the mixed-modeling group over the other two modeling groups during the first immediate retention test, whereas no differences emerged amongst the groups during the 10-minute and 24-h retention tests that occurred after the physical practice.

We speculate that the introduction of physical practice overrides the impact of the different model types gained from the observational experience. This speculation is supported by results from Andrieux and Proteau (2014). In their research, similar modeling groups as that of our design received interspersed observational and physical practice and were compared to physical practice only and control groups. No differences were shown across immediate and 24-h delayed retention tests for the different modeling groups compared to the physical practice group for those dependent measures that were similar to ours. Andrieux and Proteau, however, did show that the mixed-model and skilled observation groups were significantly better than the novice observation group for an interval timing measure that was specific to their task (and was not a component of our task). Thus, it is possible that different observation experiences target varied dimensions of a movement; e.g., temporal versus spatial components. Regardless, our evidence,

coupled with that of Andrieux and Proteau, strongly suggests that interspersing physical and observational practice may differentially impact the learning effects of different model-types as compared to the sequential use of these two practice types.

Moving onto the second objective, not surprisingly, there was no evidence to support differential consolidation processes as a result of observing different model types. The comparisons for the immediate retention tests versus the 24-h delayed test or the week-long delayed test showed no differences in performance across the modeling groups at any of the time points. Overall, the consistently better performance of the experimental groups compared to the control group across all retention intervals compared was used as an indicator that consolidation processes allowed for performance stabilization. The maintenance of performance over a retention interval greater than 24 h has been previously demonstrated both in the motor learning (Robertson et al., 2018) and consolidation literature (Karni & Sagi, 1993).

While performance stabilization was evident, the lack of improvement between the immediate and delayed retention tests showed that no offline learning occurred. Trempe et al. (2011) have suggested that offline learning may not occur if participants reach a given skill level by the end of acquisition. Therefore, it is possible that our participants reached this specific skill level before the end of their first acquisition session, given that there were no performance improvements across the latter half of the acquisition blocks, nor any performance changes seen in session 2. Perhaps if the training protocol had fewer acquisition trials before the first set of retention tests, there would have been opportunity to determine if different consolidation outcomes would occur as a function of the model type. Moreover, this quick learning of the skill resulted in our third objective, related to the impact of a longer acquisition timeframe unresolved. Further research on this topic may benefit from examining multiple acquisition sessions with fewer acquisition trials in each session.

Along the same lines, this rapid learning of the task suggests that the task was too easy to acquire and may not have had sufficient cognitive and physical demands for its purpose. Others have suggested that tasks, which lack complexity and have low cognitive demands, may not result in offline learning processes (Malangré et al., 2014). Schmidt, Erlacher, Blischke, Brueckner, and Muller (2010), for example, asked participants to learn a sequence of unrestrained arm movements to touch rectangles on a smartboard as fast as possible and, similar to our findings, they failed to see any behavioural outcomes of offline learning. Thus, perhaps the waveform-matching task was too low in complexity and not cognitively stimulating enough to see behavioural outcomes resulting from offline processes of consolidation. Consequently, continued research on this topic, but using a more difficult task, may be merited. Further, the majority of consolidation research has been done with only a limited amount of tasks, namely serial reaction time tasks (SRTTs), finger-tapping tasks, motor adaptation tasks (Malangré et al., 2014), and only a handful of studies used gross-motor skills (Schmidt, Erlacher, Blischke, Brueckner, & Müller, 2010; Trempe et al., 2011). Therefore it is possible that offline learning simply cannot be seen within this specific waveform-matching task and our results should not be generalized to other tasks.

In conclusion, we found no evidence for a differential impact of model types on the learning outcomes of an elbow flexion–extension task. Further, no differences in either performance stabilization or offline learning were evidenced, suggesting similar consolidation processes occurred across the three different model types. Despite these non-significant findings there were two points of interest. First, the temporal accuracy data at the 24-PA1R showed some support for the unique contributions of observational learning compared to physical practice alone, at least for short term retention, as the observation groups performed significantly better than the PP group. The second point of interest was that participants who had much fewer physical practice trials learned this movement, ostensibly of low task difficulty, to the same degree as those who had physical practice 100 percent of the time. This finding is meaningful for settings where continued physical practice could lead to fatigue and thus possible safety issues or stress on the body leading to injury; in such situations interspersed observational and physical practice would appear to be more effective.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humov.2018.11.014.

References

Al-Abood, S. A., Davids, K., Bennett, S. J., Ashford, D., & Marin, M. M. (2001). Effects of manipulating relative and absolute motion information during observational learning of an aiming task. *Journal of Sports Sciences, 19*(7), 507–520. https://doi.org/10.1080/026404101750238962.

Albouy, G., Fogel, S., Pottiez, H., Nguyen, V. A., Ray, L., Lungu, O., ... Doyon, J. (2013). Daytime sleep enhances consolidation of the spatial but not motoric representation of motor sequence memory. *PLoS ONE, 8* e52805 doi: 10.1371.

Andrieux, M., & Proteau, L. (2014). Mixed observation favors motor learning through better estimation of the model's performance. *Experimental Brain Research, 232*(10), 3121–3132. https://doi.org/10.1007/s00221-014-4000-3.

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature, 382*, 252–255.

Buchanan, J. J., & Dean, N. J. (2010). Specificity in practice benefits learning in novice models and variability in demonstration benefits observational practice. *Psychological Research, 74*(3), 313–326. https://doi.org/10.1007/s00426-009-0254-y.

Carter, M. J., & Ste-Marie, D. M. (2016). An interpolated activity during the knowledge-of-results delay interval eliminates the learning advantages of self-controlled feedback schedules. *Psychological Research*. https://doi.org/10.1007/s00426-016-0757-2.

Carter, M. J., & Ste-Marie, D. M. (2017). Not all choices are created equal: Task-relevant choices enhance motor learning compared to task-irrelevant choices. *Psychonomic Bulletin & Review, 24*(6), 1879–1888. https://doi.org/10.3758/s13423-017-1250-7.

Carter, M. J., Smith, V., Carlsen, A. N., & Ste-Marie, D. M. (2018). Anodal transcranial direct current stimulation over the primary motor cortex does not enhance the learning benefits of self-controlled feedback schedules. *Psychological Research, 82*(3), 496–506. https://doi.org/10.1007/s00426-017-0846-x.

Heyes, C. M., & Foster, C. L. (2002). Motor learning by observation: Evidence from a serial reaction time task. *The Quarterly Journal of Experimental Psychology, 2*, 593–607. https://doi.org/10.1080/0272498014300038.

Kantak, S. S., Sullivan, K. J., Fisher, B. E., Knowlton, B. J., & Winstein, C. J. (2010). Neural substrates of motor memory consolidation depend on practice structure. *Nature Neuroscience, 13*(8), 923–925. https://doi.org/10.1038/nn.2596.

Kantak, S. S., & Winstein, C. J. (2012). Learning–performance distinction and memory processes for motor skills: A focused review and perspective. *Behavioural Brain Research, 228*(1), 219–231. https://doi.org/10.1016/j.bbr.2011.11.028.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 363,* 210–211.

Kovacs, A. J., Boyle, J., Grutmatcher, N., & Shea, C. H. (2010). Coding of on-line and pre-planned movement sequences. *Acta Psychologica, 133*(2), 119–126. https://doi.org/10.1016/j.actpsy.2009.10. 007.

Krakauer, J., & Shadmehr, R. (2006). Consolidation of motor memory. *Trends Neruoscience, 29*(1), 58–64. https://doi.org/10.1016/j.tins.2005.10.003. CONSOLIDATION.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(4), 730–746. https://doi.org/10.1037/0278-7393.9.4.730.

Lee, T., & White, M. (1990). Influence of an unskilled Model's practice schedule on observational motor learning. *Human Movement Science, 9,* 349–367. https://doi.org/10.1016/0167-9457(90)90008-2.

Lin, C. H., Fisher, B. E., Wu, A. D., Ko, Y. A., Lee, L. Y., & Winstein, C. J. (2009). Neural correlate of the contextual interference effect in motor learning: A kinematic analysis. *Journal of Motor Behavior, 41*(3), 232–242.

Malangré, A., Leinen, P., & Blischke, K. (2014). Sleep-related offline learning in a complex arm movement sequence. *Journal of Human Kinetics, 40,* 7–20. https://doi.org/10.2478/hukin-2014-0002.

McCullagh, P., Ste-Marie, D. M., & Law, B. (2013). Modeling: Is what you see what you get? In J. Van Waalte, & B. Brewer (Eds.). *Exploring sport and exercise psychology* (pp. 139–162). ((3rd ed.).). American Psychological Association.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Science Education, 15,* 625–632. https://doi.org/10.1007/s10459-010-9222-y.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

Rohbanfard, H., & Proteau, L. (2011). Learning through observation: A combination of expert and novice models favors learning. *Experimental Brain Research, 215*, 183–197. https://doi.org/10.1007/s00221-011-2882-x.

Robertson, E. M. (2009). From creation to consolidation: A novel framework for memory processing. *PLoS Biol, 7* e19.

Robertson, E. M., Pascual-leone, A., & Miall, R. C. (2004). Current concepts in procedural consolidation. *Nature Reviews: Neuroscience, 5*.

Robertson, R., Germain, L. S., & Ste-Marie, D. M. (2018). The effects of self-observation when combined with a skilled model on the learning of gymnastics skills. *Journal of Motor Learning and Development, 1–30*. https://doi.org/10.1123/jmld.2016-0027.

Schmidt, S., Erlacher, D., Blischke, K., Brueckner, S., & Müller, F. (2010). Sleep related enhancement in motor performance: Gross vs. fine motor tasks. *Journal of Sport and Exercise Psychology, 32,* 124.

Schmidt, R. A., & Lee, T. (2011). *Motor Control and Learning: A Behavioural Emphasis* (5th ed.). Windsor, ON: Human Kinetics.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*(2), 179–187.

Shea, C. H., Wright, D. L., Wulf, G., & Whitacre, C. (2000). Physical and observational practice afford unique learning opportunities. *Journal of Motor Behavior, 32*(1), 27–36. https://doi.org/10.1080/00222890009601357.

Starek, J., & McCullagh, P. (1999). The effect of self-modeling modeling on the performance of beginning swimmers. *The Sport Psychologist, 13,* 269–287.

Ste-Marie, D. M., Law, B., Rymal, A. M., Jenny, O., Hall, C., & McCullagh, P. (2012). Observation interventions for motor skill learning and performance: An applied model for the use of observation. *International Review of Sport and Exercise Psychology, 5*(2), 145–176. https://doi.org/10.1080/1750984X.2012.66507.

Trempe, M., Sabourin, M., Rohbanfard, H., & Proteau, L. (2011). Observation learning versus physical practice leads to different consolidation outcomes in a movement timing task. *Experimental Brain Research, 209*(2), 181–192. https://doi.org/10.1007/s00221-011-2540-3.

Van Der Werf, Y. D., Van Der Helm, E., Schoonheim, M. M., Ridderikhoff, A., & Van Someren, E. J. W. (2009). Learning by observation requires an early sleep window. *Proceedings of the National Academy of Sciences, 106*(45), 18926–18930. https://doi.org/10.1073/pnas.0901320106.

van Schalkwijk, F. J., Benjamins, J. S., Migliorati, F., de Nooijer, J. A., van Someren, E. J. W., van Gog, T., & van der Werf, Y. D. (2015). The role of sleep timing in children's observational learning. *Neurobiology of Learning and Memory, 125,* 98–105. https://doi.org/10.1016/j.nlm.2015.08.003.

Winstein, C. J., & Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 677–691. https://doi.org/10.1037/0278-7393.16.4.677.

Wulf, G., Schmidt, R. A., & Deubel, H. (1993). Reduced feedback frequency enhances generalized motor program learning but not parameterization learning. *Journal of Experimental Psychology Learning Memory and Cognition, 19,* 1134.