Running head: DELIBERATE PRACTICE

http://www.ncbi.nlm.nih.gov/pubmed/24001022

How experts practice: A novel test of deliberate practice theory

Abstract

Performance improvement is thought to occur through engagement in deliberate practice. Deliberate practice is predicted to be challenging, effortful, and not inherently enjoyable. Expert and intermediate level Gaelic football players executed two types of kicks during an acquisition phase and completed pre-, post-, and retention tests. During acquisition, participants self-selected how they practiced and rated the characteristics of deliberate practice for effort and enjoyment. The expert group predominantly practiced the skill they were weaker at and improved its performance across pre-, post- and retention tests. Participants in the expert group also rated their practice as more effortful and less enjoyable compared to those in the intermediate group. In contrast, participants in the intermediate group predominantly practiced the skill they were stronger at and improved their performance from pre-test to post-test but not on the retention test. Findings provide support for deliberate practice theory and give some insight into how experts practice and possibly learn and improve their performance beyond its current level.

Keywords: Learning; Skill Acquisition; Expert Performance

How experts practice: A novel test of deliberate practice theory

An activity that is central to learning is deliberate practice. Deliberate practice is designed to improve key aspects of current performance, is challenging, effortful, requires repetition and feedback, and may not be inherently enjoyable or immediately rewarding (Ericsson, 2003; 2007; 2008). Ericsson, Krampe and Tesch-Römer (1993) provide a theoretical framework showing how deliberate practice leads to improvements in performance and the attainment of expertise. First, the "monotonic benefits assumption" (pp. 368) holds that the amount of time invested in domain-specific deliberate practice activities is positively, even monotonically, correlated to the attained performance level. Second, the individual requires resources including good teachers and suitable facilities in order to optimise practice. Third, individuals who engage in deliberate practice are predicted to rate it as more relevant to improving performance, more effortful, and less enjoyable when compared to other activities. The predictions of deliberate practice theory have typically been tested using the retrospective recall methodology in which participants are required to evaluate activities they have engaged in previously. However, ratings of practice may be confounded by a number of factors, such as lapses in memory between engaging in the practice and retrospectively rating it sometime later. To our knowledge, no researchers have previously measured the ratings of deliberate practice *during* a practice session. A novel test of deliberate practice theory is reported in this manuscript in which ratings of practice are recorded during practice itself, rather than retrospectively sometime after the practice has occurred.

Ericsson et al. (1993) used recall interviews and diaries to retrospectively examine the activities that musicians attending the West Berlin Music Academy had engaged in since starting in the domain. In their first study, violinists were divided into four groups differentiated by level of attainment. The groups were the best violinists in the Academy, good violinists, music teachers, and middle-aged professional violinists playing in world-class orchestras. The mean start age of participants in violin practice was 7.9 years of age. By 18 years of age, the best violinists and the middle-aged professional violinists had accumulated 7,410 and 7,336 hours in deliberate practice activity, respectively. In comparison, by 18 years of age the good violinists had accumulated 5,301 hours, whereas the music teachers had accumulated only 3,420 hours. The amount of deliberate practice the violinists had accumulated across their life span was monotonically related to level of attainment. In their second study, further support for this prediction was found with expert pianists having accumulated 7,606 hours of practice by 18 years of age, which was significantly more than that of amateur pianists who had accumulated only 1,606 hours. Several other researchers have subsequently provided support for the "monotonic benefits assumption" (Ericsson et al., 1993; pp. 368) across a variety of domains (e.g., Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005; Hodges & Starkes, 1996). Moreover, a recent reanalysis of the studies in chess showed that the amount of accumulated deliberate practice alone accounted for 30% of the variance in attainment level (Hambrick, Oswald, Altmann, Meinz, Gobet, & Campitelli, 2013; see also de Bruin, Smits, Rikers, & Schmidt, 2008).

The "monotonic benefits assumption" (Ericsson et al., 1993; pp. 368) only addresses a relationship between the amount of deliberate practice and attainment level. It does not address differences in the quality or efficiency of the deliberate practice engaged in, which might be expected to account for a substantial proportion of variance in eventual attainment. For example, differences in the type of deliberate practice activity engaged in have been shown to account for variation in attainment (e.g., Young & Salmela, 2010) and skill acquisition (e.g., Ford, Low, McRobert, & Williams, 2010). Ericsson et al. (1993) required violinists to rate the quality of their deliberate practice activities for "the most typical recent week" (p. 373). It was predicted that the violinists would rate deliberate practice higher than other activities for its relevance to improving performance and for the effort invested in it, but lower for enjoyment. Rating scores were based on whether they were significantly higher or lower than the grand mean for all activities. Participants were asked to view a taxonomy of activities, which was made up of 10 categories of *everyday activities* (e.g., household chores, shopping, leisure, sleep) and 12 categories of *musical activities* (e.g., solo performance, group performance, practice alone, practice with others). They were asked to rate how relevant each activity was to improving performance, how much effort was required to do the activity, and the level of enjoyment they experienced when engaging in the activity.

There were no between-group differences in the activity ratings for relevance, effort, and enjoyment. Rating scores for each activity were collapsed across groups and compared against the grand mean for all activities to determine whether they were significantly higher or lower. Sleep was the only everyday activity that scored higher for relevance than the grand mean. In terms of the musical activities, practice alone was given the highest rating for relevance to improving performance, whereas playing for fun alone was given one of the lowest ratings for relevance. The other musical activities that were rated higher for relevance than the grand mean rating of all activities were practice with others, taking lessons, solo and group performance, music theory, and listening to music. All of the musical activities that were rated higher than the grand mean for their relevance to improving performance were rated higher than the grand mean for effort and lower for enjoyment, except for listening to music and, for enjoyment only, group performance.

The relevance, effort, and enjoyment predictions outlined in the theory have been examined by only a few researchers (Helsen, Starkes & Hodges, 1998; Hodge & Deakin, 1998; Hodges & Starkes, 1996; Hyllegard & Yamamoto, 2005; Ward, Hodges, Starkes & Williams, 2007; Young & Salmela, 2002). The relevance prediction has been supported because certain domain-specific activities received higher ratings for their relevance to improving performance when compared to other activities. In sport (e.g., Hodges & Starkes, 1996), a number of domain-specific tasks have been rated by participants as higher for relevance to improving performance compared to other activities. These tasks include practice that simulates the competition environment, some aspects of physical training (e.g., weight training), practice with the coach, and sleep. However, retrospective participant ratings of the relevance of an activity to improving performance do not provide evidence that engaging in a specific activity actually led to or caused an improvement in performance.

A further prediction of deliberate practice theory is that such practice will be rated by participants as effortful. Ericsson et al. (1993) originally conceptualised effort in relation to the higher intensity and longer duration of deliberate practice compared to other activities, how these increase as the performer develops, and how it leads to the need for adequate rest and recovery. The violinists in the Ericsson et al. (1993) study provided support for the effort prediction because the musical activities they rated as higher than the grand mean for relevance to improving performance were also rated higher for effort. The prediction that deliberate practice activities will be rated higher for effort compared to other activities has been supported in a number of subsequent studies (e.g., Hodge & Deakin, 1998; Ward et al., 2007). More recently, researchers have differentiated effort into either mental or physical effort (Hodges & Starkes, 1996). In domains that require both, such as triathlon (Baker, Deakin, & Côté, 2005; Yeo & Neal, 2004), ratings for both measures are higher for deliberate practice compared to other activities.

Another prediction of deliberate practice theory is that performers will rate it as less enjoyable when compared to other domain-specific or everyday activities. Participants are presumed to engage in deliberate practice because it improves future performance, rather than for enjoyment during the activity itself (Ericsson et al., 1993). In the Ericsson et al. (1993)

study, the violinists rated the musical activities that were higher in relevance to improving performance as lower for enjoyment when compared to the grand mean of all activities. However, some researchers provided evidence contradicting the enjoyment prediction of deliberate practice theory. For example, athletes rated the activities that they had identified as being higher for relevance to improving performance (e.g., practice with the coach, games and tactics, Helsen et al., 1998; Hodges & Starkes, 1996) as *higher* for enjoyment than the grand mean.

The measures of relevance, enjoyment, and effort have not previously been recorded during practice. Participants have aggregated retrospectively their perceptions of an activity that they have engaged in many times into a single rating. Ratings that are aggregated retrospectively could create different perceptions from those actually experienced during the activity. Moreover, a number of other factors may have led athletes (e.g., Hodges & Starkes, 1996) to rate activities they identified as being highly relevant to improving performance as enjoyable. First, the social interaction and environment of sport might interfere with participant recollections of their in-the-moment enjoyment of a practice activity (Ericsson, 1996; Hodges, Kerr, Starkes, Weir & Nananidou, 2004). Second, the method of retrospectively rating the enjoyment of an activity by aggregating the perceptions of an activity that has been engaged in many times into a single rating could lead to changes in those perceptions. A superior method may be to collect ratings during or immediately after the activity (Hyllegard & Yamamoto, 2005). Third, evidence from research examining the microstructure of practice environments in sport (Deakin & Cobley, 2003; Deakin, Starkes, & Allard, 1998; Ford, Yates, & Williams, 2010; Starkes, Deakin, Allard, Hodges, & Hayes, 1996) shows that activities rated as highly relevant to improving performance are either not engaged in at all or are only engaged in for short periods. For example, Deakin et al. (1998) reported that elite figure skaters invested more practice time on jumps that they had already mastered compared to new, yet-to-be mastered, and more difficult jumps. A lot of practice time is not spent in deliberate practice and participants may have included these activities in their aggregated ratings of enjoyment.

Deliberate practice theory does not address the underlying structure of the activity being engaged in beyond the concept of repetition in practice. The underlying structure of practice has been shown to affect the amount of performance improvement or learning that occurs during practice. Random practice scheduling has generally been shown to be better for learning compared to blocked practice scheduling in a number of tasks, including a barrier knock-down task (Shea & Morgan, 1979), complex police judgments (Helsdingen, van Gog, & van Merrienboer, 2011), badminton serves (Goode & Magill, 1986), hand-writing (Ste-Marie, Clark, Findlay & Latimer, 2004), and problem solving in mathematics (Rohrer & Taylor, 2007). In a similar vein, self-selected practice schedules in which participants control the order of the practice are more effective for learning compared to schedules selected by others (e.g., a coach or experimenter), including random practice (Day, Arthur, & Gettman, 2001; Hodges, Edwards, Luttin, & Bowcock, 2011; Holladay & Quiñones, 2003; Keetch & Lee, 2007; Wulf, Raupach, & Pfeiffer, 2005). The benefits of random and possibly self-selected practice scheduling over others is thought to be caused by the performer engaging in more elaborate processing of information across each skill (Shea & Morgan, 1979) or by having to reconstruct the action plan for each skill (Lee & Magill, 1983). The additional mental effort engaged in by the learner is thought to play a key role in the learning of tasks (Lee, Swinnen, & Serrien, 1994). In support of these findings, when researchers (e.g., Baker et al., 2005) have measured mental effort invested in deliberate practice, experts have rated it as being higher than in other activities in which they engage.

We use a novel approach to examine the predictions of deliberate practice theory *during* the practice of complex, domain-specific tasks by expert and intermediate performers in Gaelic football. Gaelic football is an invasion field sport consisting of 15 players on each team who score points by passing a ball between the opposition goalposts. Expert and intermediate level Gaelic football players practiced across four sessions between pre- and post-tests to improve the performance of two kicking skills either executed from the ground or from the hands at a goal target 25 m away. Another expert group acted as controls by performing the pre- and post-tests only. During each practice session, participants were free to self-select their practice schedule and were required to rate the activity engaged in for effort and enjoyment. When participants rate their perceptions of practice during a session it is hypothesized that those ratings would accurately reflect their perceptions of that session. The participants in the expert group are hypothesized to self-select to practice the skill most relevant to their aim of improving performance, whereas the intermediate group may not. The expert group are predicted to engage in deliberate practice, whereas the intermediate group are not, or will engage in it to a lesser degree. It is hypothesized that participants in the expert group will rate their practice as more effortful and less enjoyable in comparison to the intermediate group. Finally, the expert group may be expected to self-select to execute kicks in a manner that is different from the intermediate group, perhaps through a more random as opposed to a more blocked practice schedule (e.g., Shea & Morgan, 1979), although it is possible that the participants are unaware of this principle and that there will be no between-group differences.

Method

Participants

A total of 45 male, Gaelic football players were participants. The expert group (*n* = 15; *M* age = 22.1 years, *SD* = 0.8, *M* playing years = 15.0 years, *SD* = 1.6) were contracted to senior Gaelic football teams that play at the highest level of the sport in Ireland. The intermediate group (*n* = 15; *M* age = 19.7 years, *SD* = 1.4, *M* playing years = 14.7 years, *SD* =

1.6) played lower-level amateur Gaelic football in Ireland. The control group (*n* = 15; *M* age = 22.4 years, *SD* = 1.1, *M* playing years = 15.7 years, *SD* = 1.4) were also contracted to senior Gaelic football teams. Participants provided informed consent and the research work was conducted according to the ethical guidelines of the lead institution.

Materials and Apparatus

The task required participants to execute kicks either from their hands or from the ground toward Gaelic football goalposts with the intention of getting the ball over the crossbar to score. The experimental set-up and scoring system are shown in Figure 1, whereas the two types of kicks are shown in Figure 2. Free kicks from the hands or from the ground are used frequently in Gaelic football to restart play or to attempt a score following a foul on a player. The task was created so as to simulate the participants' normal training environment. Full-size Gaelic football goalposts (height = 10 m, width = 6.5 m, crossbar height = 2.5 m) were mounted on a wall in a large gymnasium using industrial-strength tape (Rhino Gaffer Tape, Herts., UK). Between the two goalposts above the crossbar, five vertical zones of 1.3 m width were created using tape. A quantifiable graded scoring system was created using these vertical zones in order to make the task suitably challenging for the participants (Guadagnoli & Lee, 2004). Participants were awarded three points if the ball entered the centre zone, two points for the zones directly to the left or right of centre, one point for the zones at the far left or right of centre. They were awarded zero points if the ball hit the goalpost or crossbar, and minus one point if the ball went wide of the goalposts or under the crossbar.

A zone for the participant to kick from was created at a distance of 25 m directly in front of the goalpost target. The zone was a $3m^2$ piece of 28 mm synthetic grass sport surface (Tarkett Prestige XM60, Laydex, Dublin, Ireland). This surface was held in place using a total of six mats with dimensions of 200 cm x 100 cm x 4 cm (Gymnova, Leicester, UK) and industrial strength velcro (Velcro Brand, Heavy Duty Stick-On Tape, Cheshire, UK). Twenty round-shaped Gaelic footballs (O'Neill's size-5 GAA All-Ireland footballs, Belfast, Northern Ireland) were placed on the ground immediately to the right of the kicking zone. A digital video camera (3CCD Digital Video Camcorder XM2 PAL, Canon, Tokyo, Japan) was positioned directly behind the practice area and was used to record performance.

Procedure

The experiment consisted of a pre-test, acquisition phase, a post-test, and a delayed retention test. The expert and intermediate groups engaged in all the tests and phases, whereas the expert-control group completed only the pre-, post- and retention tests. Prior to the pre-test, verbal instructions were provided to each participant regarding the pre-test and experimental procedures. The pre-test occurred one week before the acquisition phase and consisted of 20 free kicks toward the goal, of which 10 were from the hands and 10 were from the ground. Kicking order was divided into four sets of five kicks. Participants were allowed four trials for familiarization prior to the pre-test. Following the pre-test, each participant was informed of his score, which was calculated as a function of a maximum of 30 points for both sets of kicks. They were instructed that the post-test and the retention test would follow the same protocol as the pre-test.

The acquisition phase consisted of four practice sessions over each of four weeks. A practice session was 15 min in duration, which was divided into 3 x 5 min bouts of kicking practice. Prior to each five minute bout of kicking practice, the participants were reminded of their pre-test scores for both kicks and informed that the goal of the practice was to improve pre-test scores (Boyce, 1992). Participants were free to self-select how they practiced during each session in terms of frequency of kicks, which kick they attempted to improve, and the order they practiced the two types of kicks. There was a two minute break between the first and second bout of practice to allow collection of the footballs. There was a seven minute break between the second and third bout to allow the participants to fill in self-report measures and to allow collection of the footballs. The type of kick and score achieved on every trial were recorded using hand notation by the lead experimenter during data collection, which was checked for accuracy against the video footage.

The ratings of deliberate practice were examined using three valid and reliable self-report measures. The task was both cognitive and physical in nature, so two self-report measures were used to test the effort prediction. First, the physical effort prediction was examined using the Rate of Perceived Exertion (RPE, Borg, 1985), which is a valid and reliable tool (Chen, Fan, & Moe, 2002) used to measure the physical effort exerted during a task. It has fifteen points that range from 6 (very, very light effort or rest; 30%) to a maximum score of 20 (exhaustion; 100%). For example, Dishman, Farquhar, and Cureton (1994) reported mean RPE scores for undergraduate student participants who rode an exercise bicycle for 20 min at power outputs increasing from 125 to 175w that ranged from a mean of 11 after 5 min (or fairly light effort; 55%) to 14 at 20 min (or somewhat hard to hard effort or a steady pace; 70%). Second, the mental effort prediction was examined using the Rating Scale of Mental Effort Scale (RSME, Zijlstra & van Dorn, 1985). It is a continuous uni-dimensional scale with eight points that range from 0 (absolutely no effort; 0%), 75 (considerable effort; 50%), to 150 (extreme effort; 100%). For example, Causer, Holmes, Smith, and Williams (2011) reported mean mental effort scores for elite shotgun shooters when skeet shooting of 77 (considerable effort; 55%). Finally, the enjoyment prediction of the theory was examined using the Physical Activity Enjoyment Scale (PACES, Kendzierski & DeCarlo, 1991), which is used to examine enjoyment levels during physical activities. It consists of 18 Likert-scaled comments relating to the current activity with 11 of the comments reversed scored. Kendzierski and DeCarlo (1991) reported mean enjoyment scores of between 65% and 70% for undergraduate student participants who rode an exercise bicycle at a comfortable pace for 20 min. We would expect enjoyment scores higher than 65% to 70% to indicate greater enjoyment and scores below to indicate lower enjoyment.

Mental effort and the nature of the cognitions generated during practice were measured using concurrent verbal reports to support the data collected with the RSME scale. Think-aloud verbal reports have been shown to be a valid and reliable method of recording thought processes (Ericsson, 2006; Ericsson & Simon, 1993; Fox, Ericsson & Best, 2011). In this study, prior to the first practice session of the acquisition phase, the two groups took part in Ericsson and Kirk's (2001) training for think-aloud concurrent verbal reports, which is based on the original instructions by Ericsson and Simon (1980; 1993). During this training, participants practiced providing verbal reports with feedback during both generic and sport-specific tasks for approximately 30 min ensuring that the criteria for giving concurrent verbal reports were attained. Participants were given a brief review of the protocol for giving verbal reports prior to each practice session. Participants were instructed to only provide concurrent verbal reports during the pre- and post-kick period. Pre-kick was defined as starting from the moment a ball was picked up to the moment before the run up to kick the ball commenced. Post-kick was defined as starting from the moment the ball hit the target zone to the moment before the next ball was picked up. Verbal reports were not collected for the period when the participants were kicking the ball as the duration of this phase was approximately 3 sec, which is too short a period to collect concurrent verbal reports (Ericsson & Simon, 1993). Pilot testing revealed the pre-kick period lasted a minimum of 30 sec and the post-kick period an average of 30 sec, which is a suitable duration for providing concurrent verbal reports (Ericsson & Simon, 1993).

A lapel microphone, telemetry radio transmitter (EW3; Sennheiser, High Wycombe, England), and telemetry radio receiver (EK100 G2; Sennheiser) were used to record the participants' verbalizations. Concurrent verbal reports were recorded during the first and third bout of each practice only. Verbal reports were not recorded during the second bout so that the impact of providing verbal reports on kicking frequency could be calculated. Participants practiced this method of verbal reporting on a minimum of four trials and a maximum of six trials before the first acquisition session.

The post-test occurred on a separate week after the last week of the acquisition phase and was the same as the pre-test. The retention test occurred on a separate week that was six weeks after the post-test. The retention test was the same as the pre-test.

Data Analysis

Pre- to post- and retention test. Accuracy scores in terms of points gained for both free kicks were calculated as a function of group and test. The pre-test scores for both the free kick from the hands and from the ground for each participant were re-categorized as their weaker and stronger kick. The kicks were re-categorized as weaker and stronger so as to test the relevance prediction of deliberate practice theory, which proposes that the activity will focus on aspects of performance that require improvement. In the three instances where a participant scored equally on both kick types during the pre-test, the number of kicks that went wide in the pre-test was used to differentiate the weaker from the stronger kick. The free kick from the hands during the pre-test was categorized as the weaker of the two kick types for eight of the expert and none of the intermediate participants. The free kick from the ground during the pre-test was categorized as the weaker of the two kick types for seven expert and fifteen intermediate participants. Accuracy scores were analyzed using a factorial ANOVA with Group (expert, intermediate, control) as the between participant factor and with Test (pre-test, post-test, retention test) and Kick (weaker, stronger) as within participants factors. All significant between-participant and interaction effects were followed up using post-hoc Tukey Honestly Significant Difference tests, whereas for significant within-participant effects the Dunn-Bonferroni adjustment calculation was used. Cohen's *f* formula was used to calculate effect size for measures involving more than two means (Cohen, 1988).

Deliberate practice data. The data from each of the three self-report scales (physical effort, mental effort, enjoyment) collected during the acquisition phase were separately calculated into single mean scores for each participant that represented the amount of that variable experienced during the acquisition phase. All scale scores were mathematically transformed into percentages to make interpretation, comparison, and plotting of data clearer. Separate independent *t*-tests were used to analyze the percentage scores from each of the three scales between the expert and intermediate groups. Correlations were conducted between the higher frequency of trials in which a group executed one of the two kick types and each of the ratings.

Concurrent verbal report statements were transcribed verbatim using natural speech and other syntactical markers. Verbal reports were put into one of three predetermined categories, namely monitoring, evaluation, and planning (Ericsson & Simon, 1993). Monitoring statements were current actions or recalled statements about current events (past/present tense). Evaluation statements were some form of positive, neutral or negative assessment of a prior statement (past/present tense). Planning statements were about future actions that will or might be executed in a future situation (future tense). The lead investigator coded and calculated the mean frequencies of statements per trial during the acquisition phase for each of the three categories as a function of the two groups. A random sample of 10% of the data was coded for reliability purposes by an independent investigator and the lead investigator separately two weeks later as per guidelines from Thomas & Nelson (2001). Inter and intra observer agreements were calculated using the equation: (agreements / (agreements + disagreements) x 100 (Thomas & Nelson, 2001). The intra- and inter-observer agreement values were 97% in each instance.

The frequency of statements in each verbal report category for the expert and intermediate groups were analyzed using separate independent *t*-tests that were adjusted using the Dunn-Bonferroni calculation. The Dunn-Bonferroni adjustment used the $p < .05$ value as the base value prior to calculation. Moreover, the frequency of trials executed by participants when verbal reports were recorded during the first and third of the three bouts of kicking practice were compared for reactivity to those executed in the second bout when no verbal reports were recorded. A paired samples *t*-test was used for this purpose to analyse the frequency of trials from the second bout compared to the frequency of the other two bouts. Further paired samples *t*-tests were used separately for each group to examine the frequency of verbal reports between the two kick types.

Acquisition phase. The frequency of kicks and number of trials for the weaker and stronger kick types was calculated for the acquisition phase, with the latter being expressed as a percentage of total kicks. An independent *t*-test was used to analyze the percentage of times the weaker kick was executed across the acquisition phase by the expert versus the intermediate group. The frequency of 5 min practice blocks in which blocked or random practice was engaged in was calculated by summing blocks as a function of group and is expressed as a percentage of the total blocks per group (n = 180 blocks). As per Shea and Morgan (1979), blocked practice was defined as occurring in a 5 min practice block in which one skill was executed repetitively throughout without a switch occurring between kicks. We further defined it as occurring in a 5 min practice block in which only one switch between kicks occurred and at least 60% of trials were on one kick with the other 40% or more on the other. Similarly, as per Shea and Morgan (1979), random practice was defined as occurring within a 5 min practice block when one kick was executed for 4 or less trials consecutively before a switch to the other occurred, and so on throughout the block, without consistent repetition of the number of trials before a switch across the block.

The alpha level required for significance for all tests was set at $p < .05$. The confidence interval level was set at 95% for all tests.

Results

Pre- to post-test accuracy

Figure 3 shows the accuracy in terms of number of points scored for the (a) weaker and (b) stronger kicks of the expert, intermediate, and expert-control groups across the pre-, post-, and retention tests. Table 1 shows the statistical results for the Group x Test x Kick factorial ANOVA on number of points scored. There were significant group, test, and kick main effects in the predicted directions with the expert groups, post and retention tests, and stronger kicks all being more accurate compared to the intermediate, pre-test, and weaker kick. There was a significant Group x Test interaction. *Post hoc* analysis showed that at pre-test there were no differences between points scored by the expert (M = 16.7 points, SD = 2.9) and expert-control groups (M = 17.8 points, SD = 3.3). However, participants in the expert group scored significantly more points in the post- (M = 19.8 points, SD = 1.9) and retention-tests (M = 19.5 points, SD = 2.1) compared to their pre-test, whereas the expert-control group did not. In addition, both expert groups scored more points than the intermediate group across all tests. There was a significant Group x Kick interaction. *Post hoc* analysis showed that the intermediate group scored fewer points with the weaker kick (M = 2.7 points, SD = 5.9) compared to the stronger kick (M = 11.9 points, SD = 4.7), whereas there was no between-kick difference in points scored for the expert groups.

There was a significant three-way Group x Test x Kick interaction. *Post hoc* analysis showed that participants in the expert group improved their scores for the weaker kick from pre- (M = 14.4 points, SD = 1.8), 95% CI [12.4, 16.5] to post-test (M = 19.9 points, SD = 2.2), 95% CI [17.6, 22.2], and maintained that improvement in retention (M = 19.4 points, SD = 2.0), 95% CI [17.4, 21.4], whereas they did not improve their accuracy for the stronger kick across tests. In contrast, *post hoc* showed that participants in the intermediate group did not significantly improve their weaker kick from pre-test to post-test or retention. The *post hoc* analysis did show that the intermediate group significantly improved their stronger kick from pre- (*M* = 8.0 points, *SD* = 3.0), 95% CI [6.8, 9.2] to post-test (*M* = 14.7 points, *SD* = 3.5), 95% CI [13.4, 15.9], but not from pre- or post-test to retention test (*M* = 12.7 points, SD = 4.2), 95% CI [-0.4, 5.3]. The expert-control group did not improve accuracy for either kick across all tests.

Acquisition phase

Deliberate practice measures. Figure 4 shows the percentage scores for each of the three self-report scales examining the rating predictions of deliberate practice theory. The expert group (*M* = 57.7%, *SD* = 3.6) rated the practice sessions as less enjoyable compared to the intermediate group (*M* = 75.8%, *SD* = 9.6), $t(28) = -6.95$, $p = .00$, $d = -2.7$, 95% CI [-23.5, -12.8]. The frequency of weaker kick trials executed by the expert group during acquisition was negatively correlated to enjoyment, r (15) = -.55, $p = .03$, but not effort, with more trials leading to lower enjoyment. The expert group (*M* = 57.9%, *SD* = 5.8; greater effort, Zijlstra & van Dorn, 1985) rated the practice sessions as greater for mental effort compared to the intermediate group (*M* = 30.7%, *SD* = 14.3; some effort, Zijlstra & van Doorn, 1985), $t(28) = 6.83$, $p = .00$, $d = 2.7$, 95% CI [19.0, 35.4]. The expert group (*M* = 58.8%, *SD* = 9.5; fairly light, Borg, 1985) rated the practice sessions as greater for physical effort compared to the intermediate group (*M* = 46.8%, *SD* = 10.7; very light or gentle walking, Borg, 1985), $t(28) = 3.24$, $p = .00$, $d = 1.2$, 95% CI [4.4, 19.5]. The frequency of stronger kick trials executed by the intermediate group was negatively correlated to physical, r (15) = -.71, $p = .00$, and mental effort, r (15) = -.61, $p = .02$, but not enjoyment, with less trials leading to greater effort. The variation within the intermediate group ratings was caused by three participants with some scores that were more similar to the expert group for enjoyment (56%, 58%, 71%), mental effort (60%, 65%, 45%), and physical effort (63%, 65%, 50%).

Figure 5 shows that the expert group (M = 3.3 statements, SD = 1.4) made a greater number of verbal report statements of thoughts per trial indicating greater mental effort compared to the intermediate group (M = 1.7 statements, SD = 0.2), $t(28)$ = 4.47, p = .01, d = 2.04, 95% CI [0.9, 2.4]. The expert group made more monitoring, $t(28)$ = 2.93, p = .00, d = 1.3, 95% CI [31.5, 177. 7] and planning statements compared to the intermediate group, $t(28)$ = 2.74, p = .01, d = 1.1, 95% CI [18.1, 125.5], but there was no between-group difference in the frequency of evaluation statements, $t(28)$ = 0.08, p = .94, d = 0.98, 95% CI [-30.2, 32.8]. Typically, participants verbalised monitoring and planning thoughts during the pre-kick period, whereas they verbalised monitoring and evaluation statements during the post-kick period. For the expert group, the frequency of statements was not different between the weaker (M = 3.4 statements, SD = 1.4) and stronger kick (M = 3.2 statements, SD = 1.7), $t(28)$ = 0.3, p > .05, d = 0.1, CI [-1.0, 1.3]. For the intermediate group, the frequency of statements for the weaker kick (M = 1.6 statements, SD = 0.4) was not different in comparison to the stronger kick (M = 1.8 statements, SD = 0.3), $t(28)$ = -1.3, p > .05, d = -0.5, CI [-0.4, 0.9]. In addition, delivering a verbal report during acquisition (M = 16.3 trials, SD = 4.0) resulted in an average of 1.2 fewer trials being executed in those bouts of kicking practice compared to the bout in which no verbal reports were delivered (M = 17.4 trials, SD = 4.5), $t(119)$ = -5.4, p = .00, d = -0.3, 95% CI [-1.6, -.7].

Practice order. The expert group (M = 43.9 kicks, SD = 8.1) executed fewer trials during practice compared to the intermediate group (M = 56.4 kicks, SD = 10.1), $t(28)$ = -3.74, p = .00, d = -1.37, 95% CI [-19.4, -5.7]. Participants in the expert group (M = 66.0%, SD = 13.3%) executed their weaker skill on a greater percentage of trials during the acquisition phase compared to the intermediate group (M = 27.0%, SD = 15.1%), $t(28)$ =

7.47, $p = .00$, $d = -2.03$, 95% CI [0.3, 0.5]. The intermediate group engaged in blocked practice in 22% of practice blocks, whereas the expert group engaged in it on 17% of blocks. In contrast, the expert group engaged in random practice in 26% of practice blocks, whereas the intermediate group engaged in it on only 3% of blocks. The other two-thirds of practice blocks (Intermediate = 134 out of 180 practice blocks; Expert = 104 blocks) did not contain our definition of random or blocked practice. Moreover, three intermediate and six expert participants did not engage in any blocked practice, whilst eleven intermediate and three expert participants did not engage in any random practice. Those practice blocks contained a hybrid version of the two in which participants executed one kick for a consecutive set of five or more trials at least once and switched between kicks at least once with more than 60% of trials on one kick. Finally, the three participants in the intermediate group who had ratings that were more similar to the expert group also had other practice variables that were similar. These participants executed their weaker skill on a greater percentage of trials compared to the intermediate group (M = 51%, SD = 5 *vs*. M = 22%, SD = 10) and improved both their pre-post scores for the weaker kick (M = 5.7 points, SD = 1.5) more so than the intermediates (M = 0.8 points, SD = 1.6), as well as the stronger kick (M = 6.0 points, SD = 1.0).

Discussion

A key prediction of Ericsson et al.'s (1993) theory of deliberate practice is that it is more relevant than other activities to improving performance. In previous research (Ericsson, et al., 1993; Helsen et al., 1998; Hodge & Deakin, 1998; Hodges & Starkes, 1996; Ward et al., 2007), participants have had to rate how relevant they perceived an activity was to improving performance by retrospectively recalling a number of practice sessions and creating an aggregate score. Although this retrospective recall method has revealed much about how experts practice, it may be that some of the activities rated as relevant to improving performance did not actually improve performance, and as such, were not deliberate practice. In this study, we have objectively shown the relevance of the practice sessions to improving performance. The practice sessions lead to improved pre- to post-test performance in the weaker kick for the expert group and stronger kick for the intermediate group, which were also the kicks they practiced most. However, only the expert group maintained that performance change from the post-test to the retention test indicating relatively permanent learning, whereas the intermediate group did not. In comparison, participants in the expert-control group who did not engage in practice did not improve their kicking accuracy across the tests.

During practice, a number of other measures that were taken demonstrated support for aspects of deliberate practice theory (Ericsson et al., 1993). Ericsson et al.'s (1993) theoretical framework holds that deliberate practice will be more effortful and less enjoyable when compared to other activities. In support of these predictions, participants in the expert group rated their practice as more physically and mentally effortful and less enjoyable compared to the ratings of the intermediate group. Their ratings for enjoyment were lower than those reported by Kendzierski and DeCarlo (1991) for participants riding an exercise bicycle for 20 min, whereas the intermediate participants' scores were higher. The expert participants' ratings for mental effort were higher than those reported by elite skeet shooters (Causer et al., 2011) and their ratings for physical effort did not differ to participants riding an exercise bicycle after 5 min (Dishman et al., 1994), whereas intermediate participant's scores were lower. Moreover, when they executed more trials on the weaker kick they rated the practice lower for enjoyment and the intermediate group rated it as more effortful. The ratings of practice for the expert group contradict the findings of those who found that deliberate practice is always rated as enjoyable by expert athletes, but support those showing that it is effortful (e.g., Helsen et al., 1998; Hodges & Starkes, 1996). Moreover, further support for the theory was provided by three participants in the intermediate group who practiced their weaker kick more often compared to the rest of their group. Similar to the expert group, they improved its accuracy and rated practice as more effortful and less enjoyable in comparison to their own group. Our data suggest that other factors may have led the athletes in previous studies (e.g., Helsen et al., 1998) to retrospectively rate deliberate practice activities as enjoyable, such as the social interaction and environment of sport (Hodges et al., 2004) or the lack of engagement in practice that is deliberate during sessions (e.g., Ford et al., 2010). The ratings provided by the expert group in the current study provide support for the idea that they were engaging in a higher quality of deliberate practice compared to the intermediate group.

The greater mental effort invested on the task by the expert group compared to the intermediate group supports deliberate practice theory (Ericsson et al., 1993). It also supports previous attempts to characterize the deliberate practice activities in which experts engage (e.g., Baker et al., 2005) and published reports (e.g., Lee et al., 1994) which suggest that the amount of mental effort invested on the task plays a key role in learning. As predicted, the expert group had a higher frequency of verbal reports on each trial when compared to the intermediate group. Participants in the expert group monitored their kicks and made plans for the next kick to a greater degree compared to the intermediate group, suggesting they used the feedback available more effectively. The verbal report data for the expert participants may explain the greater mental effort invested in the practice when compared to the intermediate participants and may be a key part of how experts practice. However, the expert and the intermediate groups did not alter the nature of their verbal reports as a function of kick type, suggesting the thought processes employed reflect a general strategy used across practice.

The expert group’s practice was more effortful and less enjoyable compared to the practice of the intermediate group, which supports deliberate practice theory (Ericsson et al., 1993). As predicted, participants in the expert group chose to practice their weaker kick significantly more compared to their stronger kick, whereas those in the intermediate group chose to practice their stronger kick more than their weaker. The latter finding may suggest that the expert group focused on improving a weakness through practice, whereas the intermediate group did not. However, the accuracy scores for the two groups suggest that this may not be the case. In the pre-test, participants in the intermediate group scored 0.8 points per trial for their stronger kick only, whereas the expert group scored an average of 1.9 points per pre-test trial for the stronger kick. Participants in the intermediate group may have chosen to practice their stronger kick because performance was relatively poor at this kick in the pre-test. However, although participants in the expert group did not attain the maximum score available on the pre-test, their score was relatively high, which probably meant they chose to practice their relatively poorer weaker kick.

The expert participants were predicted to self-select to execute the two kicks in a more random as opposed to blocked practice schedule. A random practice schedule has been shown to facilitate motor learning more so than a blocked one (Goode & Magill, 1996; Holladay & Quiñones, 2003; Rohrer & Taylor, 2007; Shea & Morgan, 1979; Ste-Marie et al., 2004), and increase mental effort (Lee et al., 1994; Lee & Magill, 1983). There was some evidence to support the prediction that the expert group would engage in more random practice compared to the intermediate group. The expert group engaged in random practice in 26% of practice blocks compared to only 3% for the intermediate group. It is possible that the more random order of kicks used by the expert group led to a relatively permanent improvement in performance and the investment of greater mental effort on task, whereas the more blocked practice schedule used by the intermediate group may have led to the lack of long-term performance improvement and lower mental effort. However, the two groups engaged in blocked practice on a comparable number of practice blocks (Intermediate = 22%; Expert =

17%) and only around a third of practice blocks contained the two different practice schedules. The majority of practice blocks contained a hybrid of the two practice schedules in which participants practiced in random blocks of trials.

The expert group invested greater physical effort, greater mental effort and rated practice activity as being less enjoyable compared to the intermediate group, which supports deliberate practice theory. Participants in the expert group engaged in fewer trials and had a more permanent improvement in performance for a more challenging skill compared to the intermediate group, suggesting that their practice was more deliberate and of a higher quality and greater efficiency. It is apparent that expert performers accumulate more hours of deliberate practice when compared to less expert counterparts (e.g., Charness et al., 2005; de Bruin et al., 2008; Ericsson et al., 1993; Hambrick et al., 2013; Hodges & Starkes, 1996) and these differences in the quality of *how* they practice may further explain why they reach a higher level of attainment compared to others. However, the approach employed in this study is descriptive in nature and care should be taken not to infer causality from the differences in practice characteristics observed between the groups. Further research is required to show which aspect of the practice engaged in by the expert group led to the performance improvement. Moreover, research is required to show whether intermediate performers who are encouraged to engage in deliberate practice or aspects of it would show a similar performance improvement to the expert group in this study.

In summary, an expert group of Gaelic football players engaged in practice that led to a relatively permanent improvement in their kicking performance. They found practice more effortful, less enjoyable, practiced a more challenging skill, appeared to use the feedback available more effectively, and used a more random order of attempts at the skills compared to an intermediate group of participants who did not improve kicking performance permanently. Our findings provide support for the theory of deliberate practice (Ericsson et al., 1993). We have shown that expert participants practice in a deliberate manner, embracing the tenets of deliberate practice theory, and that engaging in such practice appears to facilitate improvements in performance over time.

References

Baker, J., Deakin, J. M., & Côté, J. (2005). Expertise in ultra-endurance triathletes early sport involvement, training structure, and the theory of deliberate practice. *Journal of Applied Sport Psychology*, *17*, 64-78. doi:10.1080/10413200590907577

Borg, G. (1985). *An introduction to Borg's RPE Scale*. Ithaca NY: Movement Publications.

Boyce, B. A. (1992). The effects of goal proximity on skill acquisition and retention of a shooting task in a field-based setting. *Journal of Sport & Exercise Psychology*, *14*, 298-308. Retrieved from http://psycnet.apa.org/psycinfo/1993-11712-001

Causer, J., Holmes, P. S., Smith, N. C., & Williams, A. M. (2011). Anxiety, movement kinematics, and visual attention in elite-level performers. *Emotion, 11*, 595–602.

Charness, N., Tuffiash, M., Krampe, R., Reingold, E. M., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *British Journal of Psychology, 97*, 339-351.

Chen, M. J., Fan, X., & Moe, S. T. (2002). Criterion-related validity of the Borg ratings of perceived exertion scale in healthy individuals: a meta-analysis. *Journal of Sports Sciences*, *20*, 873-899. doi:10.1080/026404102320761787

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Day, E. A., Arthur Jr., W., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology*, *86*, 1022-1033. doi: 10.1037//0021-9010.86.5.1022

Deakin, J. M., & Cobley, S. (2003). A search for deliberate practice – An examination of the practice environments in figure skating and volleyball. In J. L. Starkes, & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 115-136), Champaign, IL: Human Kinetics.

Deakin, J. M., Starkes, J. L., & Allard, F. (1998). *The microstructure of practice in sport*. Ottawa, Canada: Sport Canada Technical Report.

de Bruin, A. B. H., Smits, N., Rikers, R. M. J. P., & Schmidt, H. G. (2008). Deliberate practice predicts performance over time in adolescent chess players and drop-outs: A linear mixed models analysis. *British Journal of Psychology*, 99, 473-497. doi: 10.1348/000712608X295631

Dishman, R. K., Farquhar, R. P., & Cureton, K. J. (1994). Responses to preferred intensities of exertion in men differing in activity levels. *Medicine and Science in Sports and Exercise, 26*, 783-790.

Ericsson, K. A. (Ed.) (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. Hillsdale, NJ: Lawrence Erlbaum.

Ericsson, K. A. (2003). Development of elite performance and deliberate practice: An update from the perspective of the expert performance approach. In J. L. Starkes, & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 49-84). Champaign: IL, Human Kinetics.

Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during expert's performance on representative tasks. In K. A. Ericsson, N. Charness, P. Feltovich, & R. R. Hoffman (Eds.), The *Cambridge handbook of expertise and expert performance* (pp. 223-242). Cambridge: Cambridge University Press.

Ericsson, K. A. (2007). Deliberate practice and the modifiability of body and mind: toward a science of the structure and acquisition of expert and elite performance. *International Journal of Sport Psychology*, *38*, 4-34.

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine*, *15*, 988-994. doi: 10.1111/j.1553-2712.2008.00227.x

Ericsson, K. A., and Kirk, E. (2001). *Instructions for giving retrospective verbal reports*. Department of Psychology, Florida State University. Unpublished manuscript.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 36- 406. doi: 10.1037/0033-295X.100.3.363

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215-251. doi: 10.1037/0033-295X.87.3.215

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis; Verbal reports as data, revised edition*. Cambridge, MA: Bradford Books/MIT Press.

Ford P. R., Low, J., McRobert, A. P., & Williams, A. M. (2010). Developmental activities that contribute to high or low performance by elite cricket batters when recognizing type of delivery from bowlers' advanced postural cues. *Journal of Sport & Exercise Psychology 32,* 638-654.

Ford, P. R., Yates, I., & Williams, A. M. (2010). An analysis of practice activities and instructional behaviours used by youth soccer coaches during practice: Exploring the link between science and application. *Journal of Sports Sciences*, *28*, 483-495. doi: 10.1080/02640410903582750

Fox, C. M., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive?: A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137,* 316-344. doi: 10.1037/a0021663

Goode, S., & Magill, R. A. (1986). Contextual interference effects on learning three badminton serves. *Research Quarterly for Exercise and Sport*, *57*, 308-314. Retrieved from http://eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=EJ346392&ERICExtSearch_SearchType_0=no&accno=EJ346392

Guadagnoli, M. A. & Lee, T. D. (2004). Challenge point: framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior*, 36, 212 – 224. doi: 10.3200/JMBR.36.2.212-224

Hambrick, D. Z., Oswald, F. L., Altman, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2013). Deliberate practice: Is that all it takes to become an expert? *Intelligence*. Retrieved from http://dx.doi.org/10.1016/j.intell.2013.04.001

Helsdingen, A.S., van Gog, T., & van Merrienboer, J. J. G. (2011). The effects of practice schedule on learning a complex judgment task. *Learning and Instruction, 21*, 126-136. doi:10.1016/j.learninstruc.2009.12.001

Helsen, W. F., Starkes, J. L., & Hodges, N. J. (1998). Team sports and the theory of deliberate practice. *Journal of Sport and Exercise Psychology*, *20*, 12-34. Retrieved from http://psycnet.apa.org/psycinfo/2000-16698-002

Hodge, T., & Deakin, J. (1998). Deliberate practice and expertise in the martial arts: The role of context in motor recall. *Journal of Sport and Exercise Psychology*, *20*, 260-279. Retrieved from http://psycnet.apa.org/psycinfo/1998-12094-002

Hodges, N. J., Kerr, T., Starkes, J. L., Weir, P. L., & Nananidou, A. (2004). Predicting times from deliberate practice hours for triathletes and swimmers: What, when, and where is practice important? *Journal of Experimental Psychology: Applied*, *10*, 219-237. doi: 10.1037/1076-898X.10.4.219

Hodges, N. J., Edwards, C., Luttin, S., & Bowcock, A. (2011). Learning from the experts: Gaining insights into best practice during the acquisition of three novel motor skills. *Research Quarterly for Exercise and Sport*, *82*, 178-187. doi: 10.5641/027013611X13119541883546

Hodges, N. J., & Starkes, J. L. (1996). Wrestling with the nature of expertise: A sport-specific test of Ericsson, Krampe and Tesch-Römer's (1993) theory of "deliberate practice". *International Journal of Sport Psychology*, *27*, 400-424.

Holladay, C. L., & Quiñones, M. A. (2003). Practice variability and transfer of training: The role of self-efficacy generality. *Journal of Applied Psychology*, *88*, 1094-1103. doi: 10.1037/0021-9010.88.6.1094

Hyllegard, R., & Yamamoto, M. (2005). Testing assumptions of deliberate practice theory, relevance, effort, and inherent enjoyment of practice on a novel task. *Perceptual and Motor Skills*, *101*, 283-294. doi: 10.2466/pms.101.1.283-294

Keetch, K. M., & Lee, T. D. (2007). The effect of self-regulated and experimenter-imposed practice schedules on motor learning for tasks of varying difficulty. *Research Quarterly for Exercise and Sport*, *78*, 476-486. doi: 10.5641/193250307X13082512817543

Kendzierski, D., & DeCarlo, K. J. (1991). Physical Activity Enjoyment Scale: Two Validation Studies. *Journal of Sport & Exercise Psychology*, *13*, 30-64. Retrieved from http://psycnet.apa.org/psycinfo/1991-20212-001

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*, 730-746. doi: 10.1037/0278-7393.9.4.730

Lee, T. D., Swinnen, S. P., & Serrien, D. J. (1994). Cognitive effort and motor learning. *Quest*, *46*, 328-344. Retrieved from http://www.mendeley.com/research/cognitive-effort-and-motor-learning-1/

Rohrer, D & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*, 481-498. doi: 10.1007/s11251-007-9015-8

Shea J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 179-187. Retrieved from http://psycnet.apa.org/journals/xlm/5/2/179/

Starkes, J. L., Deakin, J., Allard, F., Hodges, N. J., & Hayes, A. (1996). Deliberate practice in sports: What is it anyway? In K. A. Ericsson (Eds.), *The road to excellence: the acquisition of expert performance in the arts and sciences, sports, and games* (pp. 181-106). Maywah, NJ: Earlbaum.

Ste-Marie, D. M., Clark, S. E., Findlay, L. C., & Latimer, A. E. (2004). High levels of contextual interference enhance handwriting skill acquisition. *Journal of Motor Behavior*, *36*, 115-126. doi: 10.3200/JMBR.36.1.115-126

Thomas, J. R., & Nelson, J. K. (2001). *Research methods in physical activity*. Champaign, IL: Human Kinetics.

Ward, P., Hodges, N. J., Starkes, J. L., & Williams, A. M. (2007). The road to excellence: deliberate practice and the development of expertise. *High Ability Studies*, *18*, 119-153. doi: 10.1080/13598130701709715

Wulf, G., Raupach, M., & Pfeiffer, F. (2005). Self-controlled observational practice enhances learning. *Research Quarterly for Exercise and Sport*, *76*, 107-111. doi: 10.5641/027013605X13076330976948

Yeo, G. B., & Neal, A. (2004). A multilevel analysis of effort, practice, and performance: Effects of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology*, *89*, 231-247. doi: 10.1037/0021-9010.89.2.231

Young B. W., & Salmela, J. H. (2002). Perceptions of training and deliberate practice of middle distance runners. *International Journal of Sport Psychology, 33*, 167-181.

Young B. W., & Salmela, J. H. (2010). Examination of practice activities related to the acquisition of elite performance in Canadian middle distance running. *International Journal of Sport Psychology*, 41, 73-90.

Zijlstra, F. & Van Dorn, L. (1985). *The construction of a scale to measure perceived effort*. The Netherlands: Delft University of Technology.

Figure captions

Figure 1. Experimental set-up including graded Gaelic football goalposts through which participants could score 3pts when the ball went into the centre grid, 2pts for one grid left or right of centre, 1 pt for two grids left or right of centre, 0pts for goalpost or crossbar, -1pt for wide of goalposts or under the crossbar in the relative grid.

Figure 2. The ‘out of the hands’ kick from starting position (2a) to ball contact (2b) and the ‘off the ground’ kick from starting position (2c) to ball contact (2d).

Figure 3. Mean (SD) outcome scores for the (a) weaker and (b) stronger kicks of the expert, intermediate and expert-control groups for the pre-test, post-test and retention test.

Figure 4. Mean (SD) scores recorded during the practice sessions using the deliberate practice tenets of enjoyment (PACES), mental effort (RSME), and physical effort (RPE) for the expert and intermediate groups during the acquisition phase.

Figure 5. Mean (SD) frequency of verbal report statements for the expert and intermediate groups for monitoring, evaluation, and planning statements.

Table captions

Table 1. Results of ANOVA on number of points scored for Group (expert, intermediate, control), Test (pre-test, post-test, retention test) and Kick (weak, strong).

Figure 1

Figure 2

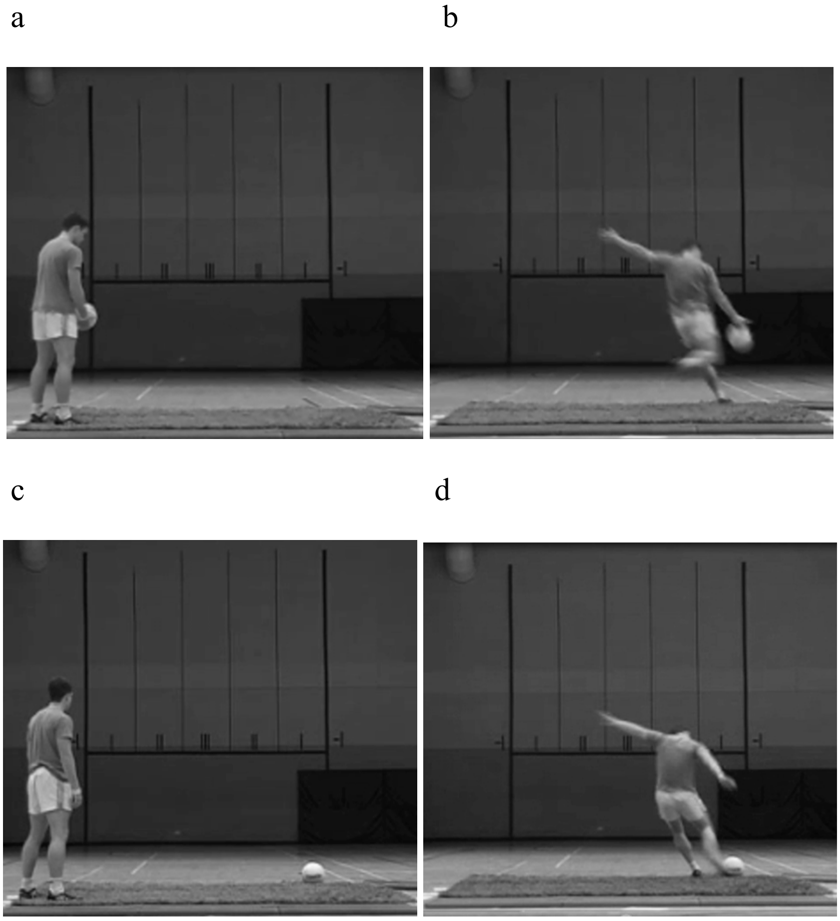
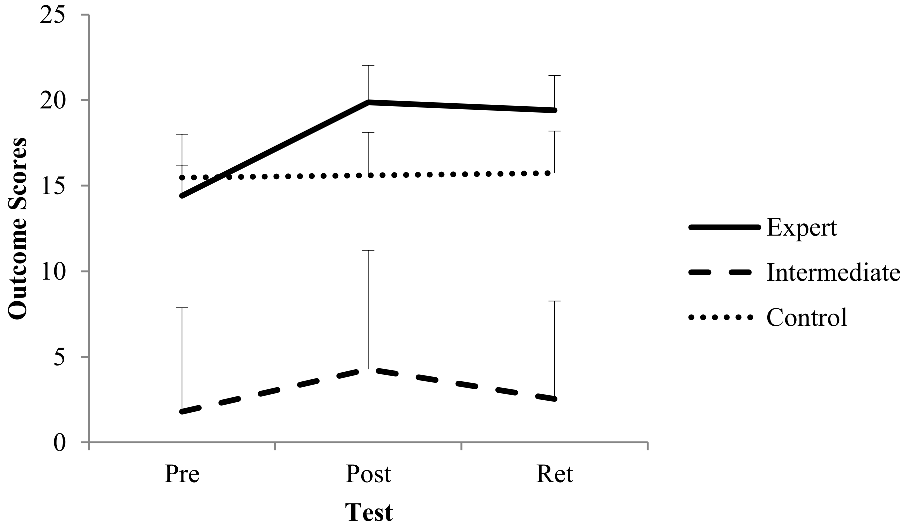


Figure 3

(a)


25
20
15
10
5
0
Outcome Scores
Pre
Post
Ret
Test
Expert
Intermediate
Control

(b)

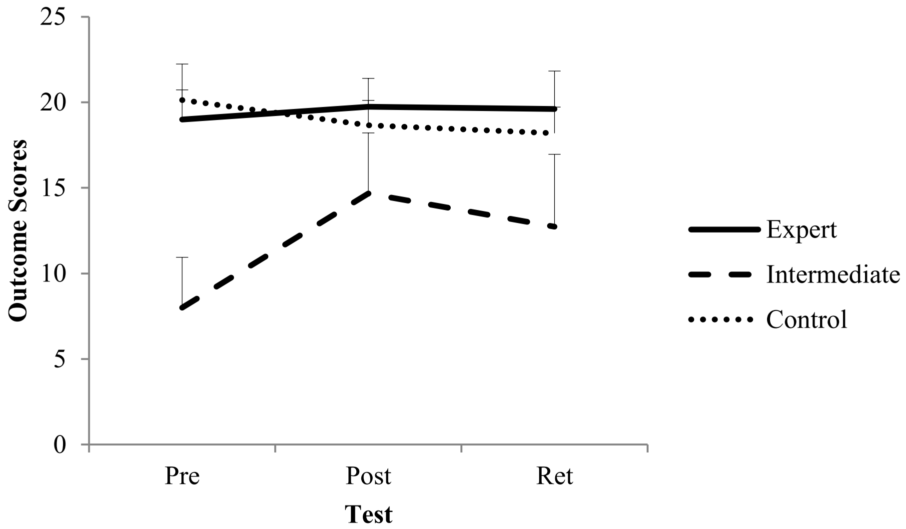

25
20
15
10
5
0
Outcome Scores
Pre
Post
Ret
Test
Expert
Intermediate
Control

Figure 4

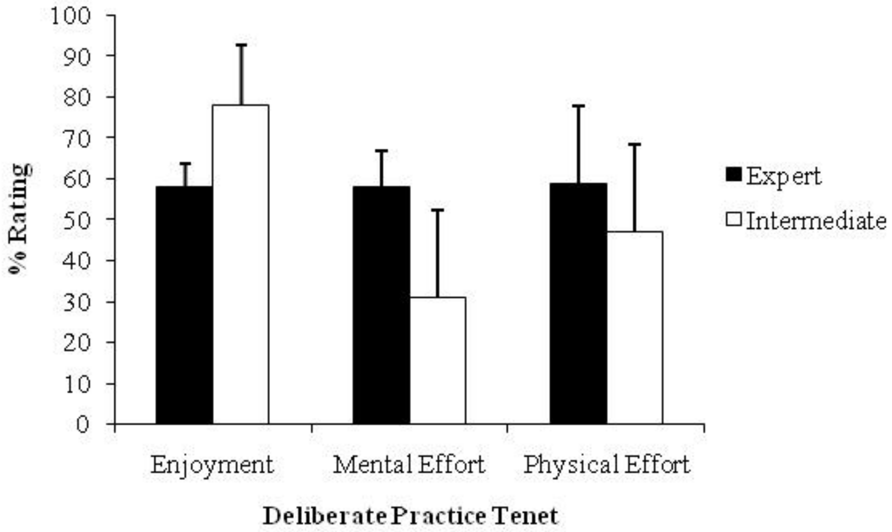


Figure 5