Metacognitive calibration in exercise: an examination using running and high-intensity functional movement exercise

Konstantinos Liverakos

A thesis submitted for the degree of PhD
at the
University of St Andrews

2021

Identifier to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/sta/86

Candidate's declaration

I, Konstantinos Liverakos, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 71,500 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2017.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 26/02/2021 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date 26/02/2021 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Konstantinos Liverakos, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis.

Printed copy

No embargo on print copy.

Electronic copy

Embargo on all of electronic copy for a period of 2 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for electronic embargo request

There are five chapters that we are preparing for publication in peer review journals. So as not to jeopardise this process we request embargo for 2 years.

Title and Abstract

- I agree to the title and abstract being published.

Date	26/02/2021	Signature of candidate
Date	26/02/2021	Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Konstantinos Liverakos, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

Date 26/02/2021 Signature of candidate

Permission for publication of underpinning research data or digital outputs

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

Embargo on all of electronic files for a period of 2 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for embargo request

We are preparing five chapters for publications in peer review journals. So as not to jeopardise this process we request embargo for two years.

Date 26/02/2021 Signature of candidate

Date 26/02/2021 Signature of supervisor

GENERAL ACKNOWLEDGMENTS

I would like to thank Elinor Bushell, Yunkai Li, Stuart McQuarrie, and Rebecca Phillips for assisting me with data collection in the Edinburgh Christmas 10k Run. Their contribution was essential in completing the study successfully. I would also like to thank Terence Crossley and Alan Roughead, the event organisers of the Edinburgh Christmas 10k Run and the Alloa Half Marathon respectively, for granting me permission to collect data from the two events. Similarly, I thank Functional Fitness St Andrews and the St Andrews Sports Centre for granting me permission to test participants in their facilities. Additionally, I would like to express my gratitude to my fellow research team researchers, Josephine Urquhart and Courtney Aitken, who were always eager to provide me with a wide range of support throughout my research project. In the same vein, I am thankful to my supervisor, Akira O'Connor, who has been heavily influential in my development as a researcher, a professional, and a person over the past five years of our collaboration. Finally, I am incredibly grateful to my partner, Elinor, and my father, Giorgos, who were always there for me to support me, encourage me, and advise me when necessary. I could not have reached this point without them.

Funding

The present work was supported by the Alfred Dunhill Links Foundation.

Research Data Access Statement

Research data underpinning this thesis are available at https://doi.org/10.17630/45fc12b1-f893-4b96-988e-6f7fec7e8d32

ABSTRACT

In the present thesis, I examined exercise calibration using running and high-intensity functional movement exercise (HIFME). In doing so, I analysed and presented results from eight studies across Chapters 2, 3, 4, and 5. In Chapters 2 and 3, I explored the extent to which demographic factors (e.g. expertise, experience, age, and gender) are effective in informing us about running and HIFME calibration. Studies 1 and 2 demonstrated that such factors do exhibit associations with running calibration, though the relatively minor strength and inconsistency of these associations also indicate that we should not overestimate the factors' contributions. Study 3 found a positive role of having a HIFME background in HIFME calibration, but other demographic factors did not exhibit associations with it. Overall, results from Chapters 2 and 3 highlighted the importance of considering demographic factors when assessing athlete calibration. However, they also highlighted the importance of understanding their limitations when doing so. In Chapter 4, I examined whether we can use self-reports of exercise metacognition and cognitive calibration to predict running and HIFME calibration. There was no significant association between any of these measures and exercise calibration in Studies 4, 5, and 6, suggesting that metacognition self-reports and calibration from other modalities are not reliable predictors of exercise calibration. In Chapter 5, I tested whether a minimal metacognitive intervention in the form of prediction guidance would lead to improved exercise calibration when participants received strategic, as opposed to impulsive, instructions. Findings from Studies 7 and 8 illustrated that strategic predictions facilitated prediction precision compared to impulsive predictions, though their effects on bias appeared to be less consistent and more dependent upon instructions. In Chapter 6, I discussed the general implications of the present thesis, and proposed ways in which future research can further explore the field of exercise calibration.

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER 1: REVIEW OF EXERCISE CALIBRATION RESEARCH

1.1 INTRODUCTION

The capacity of physical activity to confer physiological benefits has rendered it an indispensable part of the everyday life of millions of people. Such benefits include: improvements in cardiovascular fitness and body composition (Hazell, Hamilton, Olver, & Lemon, 2014; Kelley, Kelley, & Pate, 2019; Okay, Jackson, Marcinkiewicz, & Papino, 2009; Ross, Stotz, & Lam, 2015), increases in bone density (Avila, Avila, Gonçalves, & Guerra-Junior, 2019; Gabr, 2019; Sarmento et al., 2019; Suominen et al., 2017; Watson et al., 2018; Westcott, 2012), and reduction in risk of age-related disabilities (Dalgas et al., 2010; Latham, Bennett, Stretton, & Anderson, 2004; Marzetti et al., 2017), cardiovascular disease, and diabetes (Anderson et al., 2016; Fiuza-Luces et al., 2018; Pedersen, 2017; Sarmento et al., 2019; Thompson et al., 2003). In its physical exercise guidelines, the American College of Sports Medicine (ACSM) recommends that all healthy adults aged 18-65 years engage in at least thirty minutes of moderate aerobic exercise for five times per week or twenty minutes of high-intensity aerobic exercise for three times per week (Haskell et al., 2007). High- and moderate-intensity can be combined to satisfy the aforementioned criteria—high intensity refers to exercise that requires a high amount of effort and highly increases heart rate, whereas moderate intensity refers to exercise that requires a moderate amount of effort and moderately increases heart rate (World Health Organization, n.d.). The guidelines also recommend engaging in exercise that improves muscular endurance and strength (e.g. resistance training) at least twice per week. The weight of evidence is therefore such that public health organisations recommend frequent engagement in physical activity to experience health and performance (e.g. increases in strength and cardiovascular capacity) benefits.

Exercise consists of both physiological and psychological components (Avugos, Bar-Eli, Ritov, & Sher, 2013). Though people most commonly associate exercise with the former, physical activity also strongly interacts with the latter (e.g. Bashore et al., 2018; de Assis & de Almondes, 2017; Kramer & Erickson, 2007; Salam, Marcora, & Hopker, 2017). The focus of the PhD is not on the physiological effects of exercise, but on its psychological components. The present chapter is divided into two parts. In the first part (Part A), I review literature on the interactions between exercise and cognition. In the second part (Part B), I review literature on the factors that influence exercise calibration (i.e. the extent to which one's self-knowledge reflects their actual knowledge; see Section 1.2.3.2; Alexander, 2013). The first section of Part A briefly covers previous evidence on the effects of long-term exercise on cognitive functioning (e.g. executive function and memory), which is a frequently examined aspect of the relationship between cognition and exercise. The next section focuses on the influence of cognition on exercise, for which I review evidence on the cognitive demands of sports, as well as the association between cognitive fatigue, perceived exertion, and athletic performance. The final and main section of Part A explores the role of calibration in exercise. To review this relationship, I first present extensive theoretical background about metacognition (i.e. our awareness regarding our own cognition, and our control of it; Nelson & Narens, 1990) and calibration. Following this, I demonstrate the importance of calibration in exercise by discussing its role in performance, motivation, and injury risk. In Part B, which is the main focus of Chapter 1, I review evidence on the variables associated with exercise calibration (i.e. demographic characteristics and metacognition). Based on the reviewed literature, I identify factors whose associations with calibration warrant further investigation, setting up the framework for the empirical work presented in the next four chapters of the thesis. Given the scarcity of previous calibration research in exercise, I have supplemented the reviewed calibration literature with relevant pacing, weight-loss, and cognitive findings.

1.2 PART A – THE RELATIONSHIP BETWEEN COGNITION AND EXERCISE

1.2.1 Effects of long-term exercise on cognition

The effects of physical activity on cognitive functioning have garnered considerable attention, with findings indicating that individuals who exercise experience cognitive improvements (e.g. Anderson-Hanley et al., 2012; Baker et al., 2010; Cassilhas et al., 2007; Colcombe & Kramer, 2003; Liu-Ambrose et al., 2010; Northey, Cherbuin, Pumpa, Smee, & Rattray, 2018). In this section, I briefly review literature on the effects of long-term engagement in aerobic and anaerobic exercise on cognition as assessed by performance on cognitive tests and associations with neurodegenerative disease risk (e.g. dementia).

1.2.1.1 Aerobic exercise

Aerobic exercise is a very popular training method that focuses on cardiovascular conditioning and relies on aerobic metabolism (i.e. energy is generated in the presence of oxygen and without running the system into oxygen debt such that it needs to rely on anaerobic energy generation; Patel et al., 2017). Examples of aerobic exercise are long distance running, cycling, and swimming. Cross-sectional and longitudinal studies have shown that long-term aerobic exercise (one month to over a year) facilitates or prevents decline in a range of executive functions (e.g. Colcombe & Kramer, 2003; Kramer, Erickson, & Colcombe, 2006; Northey et al., 2018). Some of these functions are: task-switching (Albinet, Boucard, Bouquet, & Audiffren, 2010; Baker et al., 2010; de Assis & de Almondes, 2017; Erickson et al., 2007; Leckie et al., 2014), cognitive control (Albinet, Abou-Dest, André, & Audiffren, 2016; Anderson-Hanley et al., 2012; Baker et al., 2010; Boucard et al., 2012; de Assis & de Almondes, 2017), and working memory (Albinet et al., 2016; Erickson et al., 2007). Furthermore, Ruscheweyh and colleagues (2011) reported that low and moderate aerobic exercise improved episodic memory in older adults, whereas this improvement was not present in the sedentary control group. Overall, meta-analyses and reviews have found exercise-induced improvements in a wide range of cognitive processes and tasks (Colcombe & Kramer, 2003; Hillman, Erickson, & Kramer, 2008; Kramer et al., 2006), suggesting that exercise effects are not limited to a single process. Interestingly, long-term physical activity has also been associated with a reduction in neurodegenerative disease risk, e.g. Alzheimer's disease (Kramer & Erickson, 2007; Larson et al., 2006; Rovio et al., 2005), highlighting exercise's broad contribution to cognitive function throughout an individual's life.

1.2.1.2 Anaerobic exercise

Anaerobic exercise refers to intense and short-duration physical activity that, contrary to aerobic exercise, does not require oxygen in the mechanism of energy release, and consists of training that improves speed, muscle strength, endurance, and power (Patel et al., 2017). Examples of anaerobic exercise are weightlifting and sprinting. Mirroring aerobic exercise, anaerobic exercise is also associated with cognitive facilitation (Brown, Liu-Ambrose, Tate, & Lord, 2009; Cassilhas et al., 2007; Liu-Ambrose et al., 2010; Nagamatsu, Handy, Hsu, Voss, & Liu-Ambrose, 2012; Northey et al., 2018; Peig-Chiello, Perrig, Ehrsam, Staehelin, & Krings, 1998). Cassilhas and colleagues (2007) found that older adults engaging in 24 weeks of moderate- or high-intensity resistance training improved their performance on executive function tasks (e.g. the digit span backward task assessing working memory), whereas sedentary controls did not. Similarly, Peig-Chiello and colleagues (1998) reported that elderly adults who completed resistance training just once per week for eight weeks improved their free recall and memory recognition performance, whereas participants in the non-exercising group did not. Thus, there is considerable evidence to suggest that both aerobic and anaerobic exercise have positive effects on cognition. Accordingly, data from meta-analyses have indicated that long-term engagement in both types of exercise leads to cognitive facilitation (Colcombe & Kramer, 2003; Northey et al., 2018).

1.2.2 Influence of cognition on exercise

The relationship between cognition and exercise is bidirectional—just as physical activity can affect cognition, cognition can affect physical activity. In the section below, I briefly examine the latter relationship by reviewing literature on cognitive demands in sports, and on the effects of cognitive fatigue on physical performance.

1.2.2.1 Cognitive demands in sports

Sports often require advanced physical and cognitive skills (Schumacher, Schmidt, Wellmann, & Braumann, 2018; Wylie et al., 2018). In many sports, athletes frequently have to perform complex motor tasks in cognitively demanding situations (e.g. having to inhibit the interference effects of distractors), which could produce associations between sports engagement and cognitive ability. Evidence suggests that, compared to non-athlete controls, American football players exhibit enhanced interference and response impulse control, as well as enhanced reaction control to motion, highlighting the importance of cognitive functioning in sports (Bashore et al., 2018; Wylie et al., 2019, 2018). Similarly, metaanalytical findings indicate that athletes outperform non-athletes in attentional tasks and speed of processing measures (Voss, Kramer, Basak, Prakash, & Roberts, 2010). Interestingly, there appear to be differences in cognitive demands within sports, as player position has been associated with differences in perceptual-cognitive abilities (Bashore et al., 2018; Schumacher et al., 2018; Wylie et al., 2019, 2018), e.g. midfielders in soccer show faster acoustic reactions than defenders, and faster visual reactions than strikers. Such results show that cognitively demanding sports are associated with improved functioning in sports-relevant cognitive processes. Given the cross-sectional design implemented by studies examining cognitive performance in sports, it is not clear whether this is the result of sports-induced cognitive facilitation, or an indication of baseline cognitive contributions to sports performance, as high cognitive performers might be at an advantage compared to low performers (Bashore et al., 2018; Wylie et al., 2018). If either suggestion is true, then cognitive capacity should be taken in consideration when evaluating athletic potential in sports.

1.2.2.2 Cognitive fatigue and sports performance

Perceived exertion (also referred to as perception of effort) is an important component of physical activity, which is linked with muscle fatigue (Enoka & Stuart, 1992; Marcora, 2009). Based on perceived exertion, athletes adjust exercise duration, intensity, pace, strategy, and goals, rendering effort management essential to achieving optimal performance (Marcora, 2010; Marcora & Staiano, 2010; Pageaux, Lepers, Dietz, & Marcora, 2014). Perceived exertion has often been linked to afferent signals, i.e. input to the central nervous system from peripheral organs such as skeletal muscles and the heart (Dempsey, Amann, Romer, & Miller, 2008; Marcora, 2009; Marcora & Staiano, 2010). However, there is also evidence to support a relationship between perceived exertion and cognition, with pre-exercise cognitive fatigue (i.e. mental exertion resulting from engagement with a cognitively demanding task) leading to higher perception of effort during exercise (Marcora, Staiano, & Manning, 2009; Pageaux et al., 2014; Salam, Marcora, & Hopker, 2017; Zering, Brown, Graham, & Bray, 2017). High perception of exercise effort should in turn lead to suboptimal physical performance, illustrating that cognitive fatigue can influence athletic performance through its effect on perceived exertion.

Accordingly, empirical work has demonstrated a negative effect of cognitive fatigue on physical performance (e.g. Brown et al., 2019; Pageaux & Lepers, 2018; Van Cutsem et al., 2017). Marcora and colleagues (2009) found that, after completing a cognitively demanding task, participants gave higher ratings of perceived exertion and cycled for significantly less time before they reached exhaustion than after completing a non-demanding neutral task. Similarly, Pageaux and colleagues (2014) observed that, following a cognitively demanding inhibition task, participants were slower to complete a 5km running trial and reported higher effort than following a non-demanding task. Overall, research suggests that mental fatigue can impair athletic performance through increased perception of effort in a range of exercise activities, e.g. isometric resistance training and aerobic exercise (Brown et al., 2019; Pageaux & Lepers, 2018; Salam et al., 2017; Van Cutsem et al., 2017; Zering et al., 2017). Nonetheless, it should be noted that not all activities show impaired performance induced by cognitive fatigue, e.g. maximal anaerobic exercise is not affected by cognitive fatigue (Brown et al., 2019; Pageaux & Lepers, 2018; Van Cutsem et al., 2017). Interestingly, physiological variables typically associated with exercise, e.g. blood lactate and heart rate, have sometimes failed to mediate the effects of cognitive exertion on exercise performance (Marcora et al., 2009; Pageaux & Lepers, 2018; Pageaux et al., 2014; Van Cutsem et al., 2017). This further highlights the contribution of cognition to perception of effort and performance, suggesting that athletes should consider pre-exercise mental fatigue when they engage in most types of physical activity.

1.2.3 The role of calibration in exercise

The present section covers the main topic of Part A, which is the role of calibration in exercise. The importance of calibration and metacognition has previously been examined in learning and cognition (e.g. Gutierrez & Schraw, 2014; Kitsantas, Steen, & Huie, 2009; Schraw & Moshman, 1995; Zepeda, Richey, Ronevich, & Nokes-Malach, 2015; Zimmerman, Moylan, Hudesman, White, & Flugman, 2011). However, their role in exercise has only recently started to receive attention (Brick, MacIntyre, & Campbell, 2016; Brick, MacIntyre, & Campbell, 2014, 2015; MacIntyre, Igou, Campbell, Moran, & Matthews, 2014). Combining cognitive and exercise theoretical suggestions and findings should allow us to make inferences about the contributions of calibration to physical activity. I thus review relevant cognitive and exercise literature below that highlights the importance of calibration in exercise through its influence on athletic performance, motivation, and injury risk.

1.2.3.1 Theoretical background

Metacognition, originally defined by Flavell (1976), is a complex structure that broadly refers to one's knowledge about one's own cognition and one's control of it (Efklides, 2008; Livingston, 2003; Norman et al., 2019; Schraw & Moshman, 1995). It is typically classified into two main components: metacognitive knowledge and metacognitive regulation (Efklides, 2008, 2011; Flavell, 1979; Jacobs & Paris, 1987; Livingston, 2003; Nelson & Narens, 1990; Ozsoy, Memis, & Temur, 2009; Ozturk, 2017; Schraw & Moshman, 1995). Metacognitive knowledge refers to knowledge about one's own cognition and general cognitive function. It consists of declarative knowledge (i.e. knowledge about one's own learning abilities and the factors that influence learning), procedural knowledge (i.e. knowledge about strategies required to complete a cognitive task), and conditional knowledge (i.e. why and when to use declarative and procedural knowledge). Metacognitive regulation refers to the strategic application of metacognitive knowledge during task preparation or in response to metacognitive experiences (e.g. task-specific information, metacognitive judgments, and metacognitive feelings that contribute to affect) to control performance and achieve one's goals. It consists of predicting, planning, monitoring, and evaluating performance. Predictions are used to inform planning prior to task engagement, which then determines initial activity strategy. During the task, performance monitoring and metacognitive experiences enable the adjustment of strategy and behaviour in accordance to one's goals and current progress. Following the conclusion of the activity, post-task evaluation of the outcome and the strategy's effectiveness contributes to future performance prediction and planning. Given the high number of metacognitive components and the complexity of the network connecting them, it is useful to think of metacognition as a system (Nelson & Narens, 1990).

Flavell (1976) proposed that the active monitoring and regulation of cognitive functioning are essential components of the metacognitive system. This was further supported by Nelson and Narens (1990) who argued that monitoring and control allow for the direct flow of information between the *object* and the *meta* levels. The former refers to operations relating to current task engagement (e.g. performing a memory test). The latter refers to high-level representations (i.e. beliefs and expectations) about the object level (e.g. "I usually perform well on memory tests"). Metacognitive monitoring of the object level allows us to reflect on our performance in relation to our current goals. Depending on the outcome of our reflection, we can choose to update or maintain our task expectations and beliefs, enabling us to then control and regulate our current task strategy and performance appropriately. This feedback loop between the object and the meta levels allows for performance optimisation in the former and belief/expectation accuracy in the latter.

1.2.3.2 How to measure metacognition

We can measure metacognition using a range of offline and online methods. Offline measures refer to the retrospective collection of data on previous use of metacognition (Veenman, 2011). They are time-efficient and cost-effective (Harrison & Vallin, 2018; Schellings & Hout-Wolters, 2011; Veenman, 2011; Winne & Perry, 2000), and are frequently implemented through the administration of self-report questionnaires such as the Metacognitive Awareness Inventory (Schraw & Dennison, 1994). This allows researchers to explore a wide range of metacognition components. However, it is difficult to evaluate the extent to which self-reports reflect actual metacognitive behaviour and skills (Efklides, 2008; Harrison & Vallin, 2018; Schellings & Hout-Wolters, 2011; Tobias, Everson, & Laitusis, 1999; Veenman, 2011). Participants could attempt to present themselves in a positive light by reporting high metacognitive engagement that does not correspond to behaviour (Tobias et al., 1999; Veenman, 2011). Additionally, those who engage in metacognition consistently could perform certain metacognitive tasks automatically, thus underestimating their metacognitive regulation and knowledge, and scoring low on self-reports (Harrison & Vallin, 2018; Zepeda et al., 2015). Nonetheless, offline measures of metacognition can assist us in collecting extensive data on a range of metacognitive components in a time-efficient manner, and are thus useful in metacognitive research.

Online measures of metacognition refer to the collection of data during task engagement (Veenman, 2011). They can involve asking participants to make prospective (e.g. judgments of learning, feelings of knowing, and performance predictions) or retrospective (e.g. confidence ratings) judgments about task performance (Efklides, 2008; Hacker, Bol, & Keener, 2012). Online measures allow for the direct examination of the relationship between metacognitive judgments and performance (Norman et al., 2019). To measure this relationship, we often use calibration, which refers to the extent to which performance judgments reflect actual performance (Alexander, 2013; Lin & Zabrucky, 1998; Schraw, Kuch, & Gutierrez, 2013). Calibration carries information on metacognitive monitoring accuracy and its effectiveness in producing accurate metacognitive knowledge (Efklides, 2014; Hacker et al., 2012; Nelson & Narens, 1990; Schraw et al., 2013). In cognition, there are multiple calibration measures such as bias, sensitivity, specificity, absolute accuracy, and relative accuracy (Fleming & Lau, 2014; Hacker et al., 2012; Schraw, 2009; Schraw et al., 2013). These measures are complementary, and can be used in tandem to gain a comprehensive understanding of how calibration influences behaviour (e.g. performance and motivation), and how it is influenced by other factors (e.g. expertise, performance, and gender). We can thus measure and examine calibration through a wide range of methods, the selection of which depends on a study's aim.

Calibration methodology in exercise is more limited than it is in cognition. Performance judgments are most typically prospective (e.g. Liverakos, McIntosh, Moulin, & O'Connor, 2018), as exercise tasks provide immediate feedback, rendering it difficult to design a study where participants can make retrospective assessments without already knowing the result (Kolovelonis & Goudas, 2018). Furthermore, it might not always be feasible to conduct multiple trials—especially in a single session—because physically demanding activities (e.g. running a marathon) can be difficult to repeat in the way cognitive tasks can. In fact, even in tasks where multiple trials can be conducted (e.g. shooting free throws in basketball); it might still be preferable to collect prediction data on overall performance (e.g. "how many free throws can you make in ten attempts?") rather than on a trial-by-trial basis (e.g. "will you make the next free throw?"). This is because immediate feedback on a previous attempt can influence the prediction for the next one (Avugos et al., 2013), limiting trial independence (e.g. high accuracy in the first five free throw attempts could lead to confidence increase and thus more optimistic predictions for the remaining attempts). Consequently, participants in exercise calibration research are often just asked to make a single performance prediction (e.g. Fogarty & Else, 2005; Fogarty & Ross, 2007; Kolovelonis, 2019; Kolovelonis, Goudas, & Samara, 2020)

Certain measures of calibration (e.g. sensitivity and efficiency) require multiple trials to be valid, as single trials do not allow for the differentiation between bias and these measures (Fleming & Lau, 2014). Though we can easily achieve this in cognitive tasks, the limited capacity of exercise studies to collect predictions for multiple trials limits the extent to which such measures are valid in exercise designs. Therefore, we typically measure exercise calibration using bias and absolute accuracy (Fogarty & Else, 2005; Fogarty & Ross, 2007; Hubble & Zhao, 2016; Kolovelonis, 2019; Kolovelonis & Goudas, 2018; Kolovelonis et al., 2020; Krawczyk & Wilamowski, 2016, 2018). The former assesses prediction overconfidence and underconfidence, whilst the latter refers to prediction precision, i.e. the absolute discrepancy between prediction and performance (Schraw, 2009). These measures can be used for singular performance predictions, rendering them an effective way of assessing exercise calibration.

1.2.3.3 Calibration and self-regulation

Calibration is an important contributor to cognition and exercise. The way in which it contributes to them is best illustrated through its role in self-regulation (Efklides, 2008, 2011, 2014; Hacker et al., 2012; Stone, 2000; Zimmerman, 2000; Zimmerman et al., 2011). Self-regulation refers to the set of behaviours implemented with the aim of achieving one's goals, and emphasises the role of individuals as the agents of their own performance (Efklides, 2011; Ramdass & Zimmerman, 2008; Zimmerman, 2000). It is characterised by its cyclical nature, and consists of both domain-general and domain-specific elements (Carpenter et al., 2019; MacIntyre et al., 2014). Metacognition, cognition, volition, motivation, and affect are all determinants of self-regulation effectiveness in learning (Boekaerts, 1996). Zimmerman (2000) argued that self-regulation is composed of the forethought, performance, and self-reflection phases. In the forethought phase, self-regulated individuals set task goals based on their self-efficacy perceptions, and devise appropriate plans to achieve them. This phase is also closely linked with motivation, as self-beliefs (e.g. self-efficacy) influence volition to perform a task. Goals and plans set in the forethought phase, combined with motivation, then affect task engagement. During the performance phase, online monitoring of performance and the task environment inform the individual about their progress in relation to their goals and predictions. Based on this feedback, they can then make appropriate strategy adjustments to facilitate current performance. Finally, post-task self-reflection assists in updating forethought components (e.g. self-efficacy), which in turn affect future task engagement, planning, and performance, thus forming a self-regulation cycle.

Efklides (2011) also proposed a self-regulation model—the Metacognitive and Affective Model of Self-Regulated Learning (MASRL)—composed of the Person and the Task x Person levels. The Person level consists of a learner's stable "trait-like characteristics", which include cognitive abilities, metacognitive knowledge and regulation skills, affect, and motivation. These components interact with each other in a top-down manner to elicit motivation to engage in a task, and set task-related goals and strategies. Efklides argued that self-regulation becomes data-driven (i.e. emphasises bottom-up processes) at the Task x

Person level, where metacognitive experiences are monitored and used to adjust strategies and control performance. Based on task and monitoring outcomes, the Person level is updated, influencing future self-regulation/behaviour. Both MASRL and Zimmerman's (2000) self-regulation models emphasise the role of interactions between different self-regulation phases and components. Most importantly, they highlight the importance of metacognition and calibration for effective self-regulation, as accuracy in performance monitoring and performance awareness is instrumental for positive behavioural outcomes (Efklides, 2011, 2014; Nietfeld & Schraw, 2002; Ramdass & Zimmerman, 2008; Stone, 2000; Thiede, Anderson, & Therriault, 2003; Zimmerman, 2000; Zimmerman et al., 2011).

1.2.3.4 Calibration and performance

Individuals with inaccurate beliefs about their capacity to perform a task are more likely to make faulty predictions, set inappropriate goals, and implement ineffective strategies that contribute to suboptimal performance (Efklides, 2011, 2014; Hacker et al., 2012). Additionally, inaccurate performance monitoring and evaluation during and after a task can reduce the probability of adjusting current and future behaviour appropriately, and can also lead to further poor calibration in the future, creating a cycle of ineffective self-regulation and suboptimal performance. This can result from both overconfidence and underconfidence (Hacker et al., 2012; Ramdass & Zimmerman, 2008; Stone, 2000), as illustrated by the pacing example presented later in the section. Therefore, through self-regulation theory, we can illustrate how poor calibration can negatively influence performance by reducing an athlete's capacity to set accurate goals, make accurate predictions, come up with successful strategies and plans, and monitor and evaluate performance effectively.

Cognitive research has demonstrated a positive association between calibration and cognitive performance, i.e. good calibration is linked with high cognitive performance (Carretti, Borella, Zavagnin, & Beni, 2011; Chiu & Klassen, 2010; de Bruin & van Gog, 2012; Dunlosky & Rawson, 2012; Gutierrez & Schraw, 2015; Gutierrez de Blume, 2017; Hacker et al., 2012; Kitsantas et al., 2009; Landine & Stewart, 1998; Nietfeld, Cao, & Osborne, 2006; Ramdass & Zimmerman, 2008; Thiede et al., 2003; Zepeda et al., 2015; Zimmerman et al., 2011). Correlational studies commonly show that high performers tend to be better calibrated than low performers (e.g. Chiu & Klassen, 2010; Kruger & Dunning, 1999). However, such studies cannot establish whether good calibration leads to good performance or vice versa. It is thus important to examine whether changes in metacognition and calibration have a direct effect on performance.

Research has shown that interventions on metacognition have a facilitating effect on cognitive performance (Gutierrez & Schraw, 2015; Gutierrez de Blume, 2017; Nietfeld et al., 2006; Zimmerman et al., 2011). Such interventions often target processes that are closely linked with calibration (e.g. performance monitoring, self-reflection, self-evaluation, etc.), and, as discussed in Section 1.3.2.2, can lead to calibration improvements. It is thus possible to argue that the effects of experimental manipulations of metacognition on performance are, at least partially, explained by changes in calibration. For example, Zimmerman and colleagues (2011) found that enhancing college student ability to self-reflect on academic feedback improved calibration (by reducing overconfidence) and mathematic performance compared to students who did not receive similar training. This could mean that students who self-reflected on their performance became less overconfident, and were thus better able to select appropriate learning strategies, leading to better performance. One limitation of such studies, however, is that factors related to the intervention other than calibration (e.g. becoming better at sitting tests) could have also contributed to improved performance. This can make assessing the exact contribution of calibration difficult. Addressing this, Dunlosky and Rawson (2012) directly supported the role of calibration in performance, as they reported that individual differences in overconfidence contributed to learning and memory retention, even after controlling for baseline memory performance. Therefore, a range of evidence suggests that calibration and changes in calibration can predict cognitive performance and changes in performance respectively.

Calibration, along with metacognition and self-regulation, is also an important contributor to athletic performance (Brick, MacIntyre, et al., 2016; Brick et al., 2014, 2015; Chatzipanteli & Digelidis, 2011; Cleary, Zimmerman, & Keating, 2006; Elferink-Gemser & Hettinga, 2017; Kolovelonis, Goudas, Dermitzaki, & Kitsantas, 2013; MacIntyre et al., 2014; McCormick, Meijen, Anstiss, & Jones, 2019; Toering, Elferink-Gemser, Jordet, & Visscher, 2009). Well-calibrated athletes should be able to facilitate their athletic performance by engaging in effective self-regulation through the generation of appropriate training and competition predictions, goals, plans, and strategies. Accordingly, a small number of correlational studies have shown a positive relationship between calibration and performance, i.e. well-calibrated athletes were more likely to perform well than poorly calibrated athletes (Kolovelonis, 2019;

Kolovelonis & Goudas, 2018). Nonetheless, as with correlational designs in cognitive research, it is not clear whether calibration leads to good performance or vice versa. Manipulations that examine this relationship are thus required.

Research using metacognition/self-regulation interventions has found a positive effect on performance in physical education activities (Chatzipanteli & Digelidis, 2011; Cleary et al., 2006; Kolovelonis & Goudas, 2013; Kolovelonis et al., 2013; Kolovelonis, Goudas, Hassandra, & Dermitzaki, 2012). However, these studies can only partially inform us about the relationship between calibration and athletic performance. This is because calibration measures to examine concurrent effects of metacognition manipulations on calibration and performance have not always been included (Chatzipanteli & Digelidis, 2011; Cleary et al., 2006; Kolovelonis, Goudas, Hassandra, et al., 2012), limiting our capacity to deduce calibration contributions to performance improvements. Furthermore, studies that included calibration measures and found that self-regulation training improved performance, did not observe an effect on calibration (Kolovelonis & Goudas, 2013; Kolovelonis et al., 2013). The lack of calibration improvements could be attributed to the self-regulation manipulations implemented not targeting calibration directly (Kolovelonis, Goudas, & Dermitzaki, 2012; Kolovelonis et al., 2013). Conversely, Kolovelonis and colleagues (2020) implemented a self-regulation intervention that targeted metacognitive processes more directly, and had a positive effect on calibration, but not performance. Overall, exercise research on the effects of metacognition interventions on performance has not been clear about the role of calibration. This could be explained by the interventions used not being sufficient to improve performance and calibration in the motor tasks examined (e.g. by being too short in duration). It is likely that calibration has a positive effect on exercise performance in a similar manner to cognition, but that the limited previous research has not been able to examine the relationship effectively.

A more clear and direct way in which we can observe the effects of calibration on athletic performance is through pacing research. Pacing refers to the distribution of energy and effort across an athletic task, and is a strong contributor to athletic performance (Brick, Campbell, Metcalfe, Mair, & Macintyre, 2016; Brick, MacIntyre, et al., 2016; Elferink-Gemser & Hettinga, 2017; McCormick et al., 2019; Skorski & Abbiss, 2017; Smyth, 2018). In marathons, athletes who start the race too fast are more likely to fatigue early, slowdown in later parts of the race, and “hit the wall” (i.e. experience dramatic increases in fatigue and reductions in speed). In contrast, athletes who adopt a conservative pace are more likely to keep a slow speed throughout the race, and run faster at its end, failing to maximise their performance capacity in the process. Both pacing styles are examples of suboptimal pacing, and can consequently impede performance (Smyth, 2018). Thus, keeping an even pace throughout the race (i.e. only have small variation in pace changes) is recommended for optimal marathon performance (Díaz, Fernández-Ozcorta, & Santos-Concejero, 2018; Smyth, 2018). Positive correlations between bias and pace slowing have been observed in marathons and half-marathons (Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018; Lepers et al., 2019). Runners who were more overconfident in their performance predictions (i.e. finished the race slower than anticipated) were also more likely to start the race too fast and slow their pace in later stages of the race. These are very interesting findings, because they illustrate how calibration bias can affect athletic performance through pacing/strategy implementation. Good calibration is thus imperative for maximum strategy effectiveness and performance optimisation.

1.2.3.5 Calibration and motivation

Motivation determines the extent to which an individual attempts or maintains an activity (Lai, 2011), and is a driving force of self-regulation (Efklides, 2011, 2014; Zimmerman, 2000, 2008). Since self-regulation is an effortful process, motivation to start and maintain an activity is essential for its success. Because of this, it is important to understand the factors that contribute to motivation. Calibration is such a factor, as the extent to which an individual is aware of their performance capacity can have strong motivational implications. In the present section, I discuss the ways in which overconfidence and underconfidence affect motivation, arguing that good calibration is required for optimal motivation.

Overconfidence can have intriguing effects on motivation. There have been suggestions that overly optimistic expectations are positively associated with increased motivation to engage in an activity, as people are more likely to attempt tasks for which they have high self-efficacy (Gonida & Leondari, 2011; Schunk & DiBenedetto, 2016; Stone, 2000). For example, Paris and Oka (1986) found that students who overestimated their reading comprehension competence performed *just as well* as well-calibrated students, whilst reporting higher academic motivation. In contrast, underconfident students exhibited the lowest reading comprehension performance, and reported the lowest academic motivation. Consequently, their results suggested that high optimism has the potential to increase motivation. However, there is substantial evidence that links overconfidence with worse performance outcomes and lower expertise compared to underconfidence and good calibration (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Kruger & Dunning, 1999; Schlösser, Dunning, Johnson, & Kruger, 2013), suggesting that Paris and Oka's (1986) findings might not be representative of overconfidence. Furthermore, even if overconfidence increases the likelihood of an individual attempting to perform an activity, it is not certain that the individual will also remain motivated enough to continue the activity.

Weight-loss research, though not exactly in the domain of cognition or athletic performance, has provided support for the suggestion that overconfident individuals are more likely to attempt an activity. Rothman (2000) reviewed evidence which indicated that overweight individuals held more optimistic expectations in anticipation of a weight-loss programme than smokers did for a smoking cessation programme, leading to higher enrolment for the former than the latter (e.g. Jeffery et al., 1993; Schmid, Jeffery, & Hellerstedt, 1989). Interestingly, smokers were more likely to succeed in their programme, indicating that accurate and modest expectations are better predictors of behaviour maintenance and goal success than unrealistic optimism. Oettingen and Wadden (1991) found that, possibly resulting from high-efficacy, women with high weight-loss expectations recorded higher attendance to a weight-loss programme and higher weight-loss compared to women with low weight-loss expectations—regardless of weight-loss fantasies. However, amongst women with low expectations, having overly optimistic fantasies regarding weight-loss (e.g. imagining that losing weight will be quick and effortless) led to losing less weight compared to having negative fantasies (e.g. imagining that losing weight will be difficult). Additionally, Sbrocco, Nedegaard, Stone, and Lewis (1999) showed that participation in a moderate weight-loss programme was more likely to lead to long-term weight-loss compared to a traditional programme that set more optimistic goals. Therefore, though overconfidence can increase motivation to attempt a task, it appears to have the opposite effect when it comes to behaviour maintenance and goal achievement.

To better understand how overconfidence affects exercise motivation, it is important to examine why it hinders activity maintenance despite its capacity to increase motivation for initial engagement. Based on the Decision Affect Theory (which predicts that surprising outcomes affect pleasure derived from a task), we would expect overconfidence to reduce pleasure derived from success, and increase failure-related disappointment (McGraw, Mellers, & Ritov, 2004; Mellers, 2000; Mellers, Schwartz, Ho, & Ritov, 1997; Mellers, Schwartz, & Ritov, 1999; Shepperd & Mcnulty, 2002). McGraw and colleagues (2004) tested this hypothesis by assessing the emotional reactions towards success and failure of recreational basketball players in a shooting task. In their first study, they recruited forty-five undergraduates who played basketball and investigated how their confidence in making a shot related to their emotions about the outcome. Athletes with higher confidence rated successful shots as less pleasant than underconfident athletes, and missed shots as more unpleasant. The authors used predictions made by the decision affect theory to calculate the pleasure scores that participants would be expected to experience with perfect calibration. They compared these scores with those reported by participants, and found that better calibration would have resulted in higher pleasure ratings. To test this suggestion directly, they recruited forty-two basketball players and allocated them in a debiased and a control group. They performed debiasing by informing participants of a general tendency to overestimate expected performance in basketball. The debiased group was better calibrated than the control group, and derived greater pleasure from the task, which resulted from less dissatisfaction over failed performance.

Similarly, Foster, Wadden, Vogt, and Brewer (1997) reported that obese women with unrealistic expectations of weight-loss in a 48-week programme rated a potential 25kg weight-loss as acceptable and a 17kg loss as disappointing. The average weight-loss recorded after the programme was 16kg, with 47% of the women failing to reach even their "disappointing" target. Furthermore, a larger discrepancy between baseline goal weights and post-intervention weights predicted lower reports of satisfaction with the programme's outcome. Consequently, reducing high overconfidence is essential in maintaining behaviour as, despite its potential to increase the likelihood of initial task engagement, it can reduce perceived achievement and pleasure derived from the activity. Making a task less enjoyable has negative implications for intrinsic motivation, i.e. motivation based on whether an activity is perceived to be enjoyable, pleasant, or interesting (Lai, 2011). According to Self-Determination Theory, intrinsic motivation is a strong and autonomous form of motivation (Deci & Ryan, 2008; Ryan & Deci, 2000; Ryan, Patrick, Deci, & Williams, 2008), and higher intrinsic motivation predicts higher exercise participation (Sibley & Bergman, 2018; Teixeira, Carraça, Markland, Silva, & Ryan, 2012). Thus, it is important to reduce overconfidence to avoid decreasing intrinsic motivation, and thus activity maintenance.

Underconfidence can also influence exercise motivation. Though underconfident individuals can be high performers and exhibit high expertise (Chiu & Klassen, 2010; Dunning et al., 2003; Kruger & Dunning, 1999; Schlösser et al., 2013), underconfidence can still have a negative contribution to exercise motivation. Contrary to overconfident athletes who might set overly optimistic goals and attempt tasks that are difficult or impossible to achieve, underconfident athletes might set overly conservative goals and avoid engaging in achievable tasks because of perceived lack of ability and anticipation of failure (Gonida & Leondari, 2011; Kolovelonis et al., 2012; Lirgg, 1991; Ramdass & Zimmerman, 2008; Schunk & DiBenedetto, 2016; Weinberg, 2009). This can have a debilitating effect on starting and maintaining a physical activity. Individuals who do not perceive themselves to be fit might avoid taking up sports they think are too advanced for them. Additionally, underconfident athletes who already participate in a sport might avoid new and challenging tasks, thereby hindering their progress. In such cases, training can be inefficient by spending time and effort on skills that have already been mastered instead of advancing to more appropriate tasks (Hacker et al., 2012). Finally, it has been suggested that completing very easy tasks provides limited pleasure (McGraw et al., 2004), so avoiding challenging tasks could also reduce intrinsic motivation. Therefore, underconfidence can have negative implications for exercise engagement and maintenance and needs to be addressed by coaches and fitness instructors. Nonetheless, it should be noted that these suggestions, although intuitive in nature, have not been tested experimentally. To my knowledge, no study has directly examined the effects of underconfidence on exercise motivation. Consequently, though we can argue that underconfidence reduces exercise motivation by limiting performance potential and reducing pleasure derived from an activity, further research is warranted to test these suggestions.

1.2.3.6 Calibration and injury risk

Just as calibration can affect athletic performance and motivation, it can also contribute to injury risk through workload selection. Training and competition loads have consistently exhibited associations with exercise injury risk (Drew, Finch, & Drew, 2016; Eckard, Padua, Hearn, Pexa, & Frank, 2018; Gabbett & Domrow, 2007; Gabbett & Jenkins, 2011; C. M. Jones, Griffiths, & Mellalieu, 2017; Malone, Hughes, Doran, Collins, & Gabbett, 2019; Rogalski, Dawson, Heasman, & Gabbett, 2013). This is because athletes who engage in high training and competition loads are more exposed to external risk factors (e.g. sports equipment malfunction risk and opponent behaviour) that increase injury susceptibility through increased exposure to inciting events (i.e. events which induce biomechanical stress that exceeds an athlete's tissue tolerance) that cause injuries (Windt & Gabbett, 2016). Additionally, high workloads can increase musculoskeletal fatigue, which contributes to internal risk factors through the reduction of neuromuscular control and tissue resilience, thereby increasing injury risk (Jones et al., 2017; Windt & Gabbett, 2016). Conversely, moderate training loads, i.e. workloads that are sufficient to improve strength and fitness but not high enough to lead to exhaustion, have a protective effect against injuries, as high strength and fitness contribute to reduced injury risk (Drew et al., 2016; Eckard et al., 2018; Jones et al., 2017; Malone et al., 2019; Windt & Gabbett, 2016). Low training loads do not improve fitness and strength enough to reduce injury risk—though, unlike high workloads, they rarely expose athletes to internal and external risk factors, or inciting events (Windt & Gabbett, 2016). Therefore, training and competition loads have a substantial contribution to injury risk, either by protecting athletes against injuries, or by increasing their susceptibility to them.

Calibration can have important implications for training and competition load selection. Overconfidence has the potential to influence both training and competition loads, with overconfident athletes being more likely to set overly optimistic goals for a competition, which can in turn differentially affect training load. On the one hand, overconfident athletes might erroneously believe they can already perform at a high level, thus failing to prepare sufficiently for the competition. In this case, they would be expected to engage in low training loads, reducing the capacity of their musculoskeletal system to handle high competition loads (Drew et al., 2016; Eckard et al., 2018; Jones et al., 2017; Rogalski et al., 2013; Windt & Gabbett, 2016). This would then render them vulnerable to injuries during the competition—especially if their competition load heavily surpasses their training loads. On the other hand, overconfident athletes who erroneously believe they can improve their fitness in a short period by engaging in very high training loads are likely to fatigue themselves, increasing their exposure to internal and external risk factors, as well as injury inciting events, during training. In this scenario, it is possible that overconfident athletes end up getting injured during the training period even before the competition. In either case, and regardless of training load approach, overconfident athletes are likely to be more susceptible to behaviour that increases injury risk during training sessions and/or competitions.

Underconfidence can also contribute to injury risk. As discussed in Section 1.2.3.5, underconfident athletes might be less likely to engage in challenging tasks (Kolovelonis, Goudas, & Dermitzaki, 2012; Lirgg, 1991; Ramdass & Zimmerman, 2008; Schunk & DiBenedetto, 2016; Weinberg, 2009), thus choosing low training and competition workloads (e.g. slow pacing strategies). Contrary to high workloads, low workloads are unlikely to expose athletes to internal and external risk factors, and injury inciting events (Windt & Gabbett, 2016). However, athletes who engage in low training loads are also less likely to improve their fitness and strength sufficiently to protect themselves from injuries (Drew et al., 2016; Eckard et al., 2018; Jones et al., 2017; Malone et al., 2019; Windt & Gabbett, 2016). Consequently, underconfidence has the potential to increase injury risk by decreasing the likelihood of athletes engaging in exercise that is intense and challenging enough to increase their resistance to injuries. In contrast, well-calibrated athletes who set realistic performance goals should be better able to select appropriate moderate workloads that improve their physical capacities enough to protect them from injuries, without risking exhaustion and high exposure to inciting events. In conclusion, calibration has the capacity to contribute to injury risk through its role in selecting appropriate (or inappropriate) training and competition loads, further highlighting its important role in exercise. Nonetheless, the relationship between calibration and injury risk has not been tested experimentally, and thus remains speculative and in need of further research.

1.2.3.7 Summary of calibration contributions to physical exercise

In Section 1.2.3, I reviewed the contributions of calibration to physical exercise by presenting and discussing its impact on performance, motivation, and injury risk. Good calibration has important implications for optimising performance, increasing motivation to start and maintain an activity, and minimising injury risk. In contrast, poor calibration is likely to contribute to suboptimal performance, low motivation, and high injury risk. Because of this, it is essential to explore and understand the factors that influence calibration, and whether it is possible to develop interventions that can improve it, and thus facilitate athletic performance, increase exercise motivation, and reduce injury risk.

1.3 PART B – FACTOR ASSOCIATIONS WITH CALIBRATION

The second and main part of the review (Part B) focused on research that has explored variables associated with exercise calibration. As discussed, calibration has important implications for athletic performance, motivation, and injury risk. However, individuals often exhibit poor calibration in a wide range of cognitive and exercise activities, with overconfidence being a common finding (e.g. Dunlosky & Rawson, 2012; Fogarty & Ross, 2007; Kolovelonis, 2019; Kolovelonis & Goudas, 2018; Lundeberg, Fox, & Punćochaŕ, 1994). Given the potential consequences of miscalibration, it is important to identify factors that contribute to good and poor calibration, and devise ways to increase the former and limit the latter. In the following sections, I review literature on the contribution of two different categories of factors to calibration. First, I focus on demographic factors (i.e. expertise, experience, age, and gender), since they can be used to diagnose calibration tendencies (e.g. a coach can predict if an athlete will be well or poorly calibrated based on previous activity experience). Second, I examine research on the influence of metacognition on calibration. Given the close relationship between metacognition and calibration (as calibration is a measure of metacognitive monitoring), we would expect improvements in metacognition to also reflect or lead to improvements in calibration. It should be noted that calibration research in exercise is still limited, so I have supplemented the topics reviewed with relevant cognitive and pacing literature where appropriate. It has been suggested that metacognitive components can apply to activities across domains (Arbuzova et al., 2020; Carpenter et al., 2019; Jonker, Elferink-Gemser, Toering, Lyons, & Visscher, 2010; Jonker, Elferink-Gemser, & Visscher, 2011), so research in cognition and learning should also inform research in exercise. Pacing research was also assumed to provide insight to calibration, as bias has been shown to correlate with pacing in running (Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018; Lepers et al., 2019).

1.3.1 Demographic factors and calibration

Demographic factors refer to individual characteristics, such as expertise, experience, age, and gender. Potential associations between demographic factors and exercise calibration would allow coaches and instructors to make inferences about an athlete's calibration in a quick and time-efficient manner, as demographic information is easy to collect. Based on these inferences, coaches, organisers of mass participation events, fitness class instructors, and health intervention promoters, among others, could then predict whether an athlete will be well or poorly calibrated, and adjust their approach appropriately. For example, a poorly calibrated athlete might need to receive performance-monitoring training tailored to their calibration needs. Conversely, a well-calibrated athlete might not need to receive any monitoring training, but could instead assist in improving the calibration of other athletes (e.g. by providing advice). In the present section, I review research on the relationships between previously researched demographic factors (expertise, experience, age, and gender) and calibration. Results on these relationships have important theoretical and practical implications for calibration research and athlete coaching.

1.3.1.1 Expertise

Expertise plays an important role in calibration. As discussed in Section 1.2.3.4, calibration contributes to cognitive and exercise performance, i.e. good calibration can facilitate performance. However, this relationship is bidirectional; expertise can also influence calibration. Individuals with high task expertise (i.e. better performance) are more likely to make accurate performance estimates than non-experts (Dunning et al., 2003; Kruger & Dunning, 1999; Schlösser et al., 2013). Low performers tend to be more overconfident and less precise than high performers. High performers are precise in their self-assessments, but also exhibit slight underconfidence—though this could be partially attributed to statistical artefacts (Schlösser et al., 2013). The tendency of low and high performers to overestimate and underestimate their abilities relative to their performance is known as the Dunning-Kruger effect, and has gathered considerable empirical support in cognitive research (Chiu & Klassen, 2010; Dunning et al., 2003; Horgan, 1992; Kruger & Dunning, 1999; Mahmood, 2016; Pajares & Graham, 1999; Schlösser et al., 2013).

In their large cohort of 88,590 students from 34 countries, Chiu and Klassen (2010) found that students who overestimated their mathematical capacity were more likely to score below the country mean, whereas underconfident students were more likely to score above it. Horgan (1992) observed that chess players under the age of sixteen with a high Elo rating (Elo rating is a measure of ability in chess, with higher ratings indicating higher expertise) were better calibrated when assessing performance probabilities for different chess scenarios than players with lower Elo ratings, who were overconfident. Additionally, Kruger and Dunning (1999) directly manipulated logical reasoning expertise by providing an experimental group with a training intervention aiming to increase logical reasoning skills, whilst a control group completed an unrelated task. Participants from both groups completed the logical reasoning task and rated their performance and competence. They then completed the experimental or control tasks, and were asked to rate their performance and skills again. Pre-training results exhibited the expected Dunning-Kruger effect (see description above), but post-training results showed that participants who received logical reasoning training improved their calibration, whereas the control group did not. More specifically, low performers in the training group reduced their overconfidence, whilst high performers reduced their underconfidence. Consequently, there is substantial evidence to exhibit a positive association of expertise with calibration in cognition. We could at least partly attribute this to experts making higher or better use of metacognition than non-experts, contributing to both good performance and high performance awareness (Dunning et al., 2003; Kruger & Dunning, 1999; Schlösser et al., 2013).

In physical exercise, the relationship between expertise and calibration has produced similar, but less consistent, results (Fogarty & Else, 2005; Fogarty & Ross, 2007; Hubble & Zhao, 2016; Kolovelonis, 2019; Kolovelonis & Goudas, 2018; Krawczyk & Wilamowski, 2016). Krawczyk and Wilamowski (2016) found that runners who were faster to complete the first half of a marathon race, i.e. had higher running expertise, were less likely to be overconfident than slower runners. Hubble and Zhao (2016) examined marathon data where they split runners into corrals based on their expected finish time and previous performance. They expected runners in faster corrals to be better calibrated than runners in slower corrals. Surprisingly, the slowest corral exhibited the least overconfidence, going against the expected expertise influence. However, the race had a time limit of six hours, meaning that runners who did not manage to finish the race in less than six hours did not have their data included in the study. Given that these runners were most likely to be in the slowest corral, this could have led to an overestimation of the corral's calibration. A comparison between the other two corrals showed that runners in the faster corral were less overconfident in their predictions than runners in the slower corral, indicating a small expertise influence on calibration.

Pacing research has been more consistent in illustrating expertise effects, with faster runners being more likely to keep an even pace throughout a marathon than slower runners, who tend to slow more in later parts of the race (Breen, Norris, Healy, & Anderson, 2018; March, Vanderburgh, TitleBaum, & Hoops, 2011; Nikolaidis & Knechtle, 2017, 2018b). Given the correlation between pace slowing and calibration bias (Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018; Lepers et al., 2019), pacing findings suggest that we should also expect faster runners to make more accurate and less overconfident predictions than slower runners. Collectively, running calibration and pacing findings are in line with cognitive literature and illustrate that expert runners are more likely to be well-calibrated than non-expert runners.

Research on the association between expertise and calibration in exercise modalities other than running has been limited (Fogarty & Else, 2005; Fogarty & Ross, 2007; Kolovelonis, 2019; Kolovelonis & Goudas, 2018). Fogarty and Else (2005) investigated whether golf handicap in fifty-four golfers predicted calibration in a (difficult) chipping task and a (relatively easy) putting task. Golfers with a low handicap had higher expertise and were thus expected to be better calibrated than golfers with a high handicap. Surprisingly, low and high expertise groups did not differ in their calibration for either task, indicating a lack of expertise influence on calibration. Fogarty and Ross (2007) recruited sixty-four tennis players and examined whether experts (i.e. former or current professional players) were better calibrated in two serving tasks of varying difficulty (i.e. one was more difficult than the other) than non-experts (i.e. junior and social players). Both expert and non-expert players exhibited similar calibration for the easier serving task, indicating no expertise influence. In the second and more difficult task however, experts were better calibrated than non-experts, as their predictions better reflected their subsequent performance. In physical education studies, Kolovelonis and Goudas (2018) and Kolovelonis (2019) found a positive relationship between expertise and basketball shooting calibration in 10-12-year-old students. Participants who scored more shots were less likely to be overconfident and more likely to be precise in their performance predictions. Overall, findings from sports other than running also indicate that high expertise is associated with better calibration. Nonetheless, this pattern of results was not present in all tasks examined, and thus the extent to which it generalises across activities needs to be further tested.

In conclusion, exercise research indicates a positive relationship between expertise and calibration, though these findings seem to be less reliable in exercise than in cognition. A reason for this could be that expertise in cognitive activities is mostly shaped by cognitive and metacognitive abilities, which also contribute to calibration (Dunning et al., 2003; Kruger & Dunning, 1999). Despite these components playing an important role in athletic performance (see Section 1.2.3.4), sports expertise is heavily determined by physical factors (e.g. strength and fitness). Because of this, individuals can have high metacognitive awareness of exercise performance and strategies without possessing the physical abilities required to be highly competent in it (Kruger & Dunning, 1999). For example, coaches and older athletes might be very knowledgeable and experienced in a sport, without being expert performers (see Section 1.3.1.3 for a discussion on the relationship between age and experience). In such cases, we would expect these individuals to be well-calibrated in their performance judgments, even if they are unable to perform at a high level. When this happens, and expertise is operationalised in terms of performance alone, it might be difficult to observe calibration differences between experts and non-experts. It thus becomes important to also consider other demographic factors, such as experience, that could contribute to calibration (Fogarty & Else, 2005).

1.3.1.2 Experience

Experience is a commonly examined demographic factor in both cognitive and exercise calibration research. Cognitive research has consistently demonstrated positive experience effects on calibration across a range of tasks and processes (Brown, Smiley, & Lawton, 1978; Carpenter et al., 2019; Krätzig & Arbuthnott, 2009; Nederhand, Tabbers, Splinter, & Rikers, 2018; Nietfeld et al., 2006; Urban & Urban, 2018). Krätzig and Arbuthnott (2009) examined the effects of item-specific experience, assessed using task repetition, on relative metamemory calibration accuracy (i.e. correlations between metacognitive judgments and performance). Greater task exposure had a facilitating effect on metamemory calibration, leading to higher positive correlations between judgments of learning and performance. This result was especially prominent for difficult tasks, possibly because easier tasks had higher baseline calibration. Similarly, Nederhand and colleagues (2018) found that medical specialists, who had more years of diagnostic experience than medical students, were better calibrated (in terms of absolute accuracy) when rating their own diagnostic performance than the less experienced medical students. Furthermore, feedback on diagnostic accuracy improved calibration across groups, suggesting that, to improve calibration, individuals need to have experience with both performing the task and evaluating their performance. Overall, cognitive research has produced results that demonstrate the positive contribution of experience to calibration. Numerous variables, such as length of activity exposure and feedback provision, can contribute to the development of task experience, which can in turn lead to more accurate performance judgments.

The relationship between experience and exercise calibration in exercise has been examined across a range of physical activities of differing complexity; e.g. running, tennis, golf, and physical education (Fogarty & Else, 2005; Fogarty & Ross, 2007; Franklin, Forgac, & Hellerstein, 1978; Kolovelonis, 2019; Liverakos et al., 2018). In running, Franklin and colleagues (1978) investigated marathon calibration in first-time, second-time, and experienced (i.e. two or more marathons completed) marathon runners. They found that experienced runners were more accurate in their predictions than first-time runners. Similarly, Liverakos and colleagues (2018) collected prediction and finish time data from a half marathon over a number of years. They operationalised experience in terms of club membership and race repetition. Club members were assumed to have higher experience compared to unaffiliated runners because of feedback and coaching availability. Race repetition was assessed longitudinally by analysing the calibration data of runners who had participated in the same race on multiple occasions. Supporting the positive influence of experience on calibration, the authors found that club runners were more accurate than unaffiliated runners. Furthermore, runners became more accurate in their third race compared to their first two. Consequently, the reviewed running studies suggest that different markers of experience can predict running calibration.

The association between experience and calibration has also been examined in exercise modalities where performance might be more difficult to predict. Running can be a challenging and technical sport, but it only consists of one movement pattern. It may thus be easier to estimate running performance compared to performance in sports that consist of numerous movements and activity patterns, and are thus more complex (e.g. basketball, tennis, and golf).[1] It is therefore important to investigate whether the influence of experience on running calibration would also be present in other, more complex sports. Kolovelonis (2019) examined the influence of extracurricular sports experience on calibration for a basketball-shooting task in 10-12-year-old students. He found that students who participated in sports outside of school were less overconfident and more precise in their shooting performance predictions than students who did not. Furthermore, more years/months of experience predicted higher precision in the experienced group. These results are interesting as they show that general sports experience is positively associated with calibration, i.e. more

[1] This is referring to sports movement patterns and skills alone, and not processes that relate to predicting and assessing opponent or teammate behaviour. Such processes can render estimating performance even more complex, and are thus outside the scope of the present thesis.

experience is linked with better calibration, even if this experience is not necessarily specific to the exercise modality in question.

Manipulating exercise experience to improve calibration can be challenging. In the study of Fogarty and Ross (2007), tennis players were asked to make performance predictions for two tennis serving tasks (an easy and a difficult one), which they then completed. This process took place twice. The initial completion of each task provided participants with performance feedback. The tennis players were then asked to complete the same tasks again. It was expected that experience derived from feedback would lead to predictions that were more accurate in the second attempt of each task. Surprisingly, feedback only improved calibration in the easy, but not the difficult, task. Therefore, experience in the form of feedback had a smaller effect on calibration than anticipated. Similarly, Fogarty and Else (2005) examined whether feedback from one task repetition would improve calibration in putting and chipping golf tasks. They found that experience attained through feedback did not improve calibration in the high- and moderate-skilled groups, with participants exhibiting similar overconfidence across task repetitions. Only participants in the low expertise group showed a significant calibration improvement for the chipping, but not the putting, task. Results from these studies suggest that, though feedback can have a positive influence on calibration, the magnitude of its contribution is likely to be small. It has been argued that, to improve calibration, feedback should not merely provide athletes with performance information, as an emphasis on calibration is also required (Kolovelonis et al., 2013). To increase exercise experience in a way that facilitates calibration, it might thus be important to provide athletes with extensive feedback that informs them about both their performance and the effectiveness of the processes they utilised to make their predictions (e.g. inform an athlete that they were overconfident and that they need to make predictions that are more conservative).

Overall, research has exhibited a positive role of experience in calibration for both cognitive and exercise domains. Nonetheless, experience is a multifaceted factor that stems from numerous variables. Because of this, it is important to explore whether and how various experience markers are associated with calibration. The reviewed research has only examined a limited number of such markers, and not always within the same exercise modality. For example, though training factors such as years of running, training volume and distance, and number of races completed before have been identified as contributors to marathon pacing (Deaner, Carter, Joyner, & Hunter, 2014; Swain, Biggins, & Gordon, 2019), their relationship with running calibration has yet to be tested. Furthermore, the association between experience and calibration (and pacing) has not always been consistent (Carlsson, Assarsson, & Carlsson, 2016; Deaner, Addona, & Hanley, 2019; Fogarty & Else, 2005; Fogarty & Ross, 2007), and, where it has been documented, the effect sizes tend to be small (Deaner et al., 2014; Kolovelonis, 2019; Liverakos et al., 2018; Swain et al., 2019). Consequently, to assess athlete calibration accurately, we need to investigate the contributions of as many experience markers as possible. Finally, it is important to test whether the relationship between experience and calibration identified in one exercise modality generalises to other modalities to produce a detailed and comprehensive account of how experience can influence calibration across different types of exercise. Therefore, we need to conduct more empirical work to explore associations between various experience markers and calibration across exercise modalities with different characteristics (e.g. complexity and predictability).

1.3.1.3 Age

Age is another demographic factor with potential implications for calibration. Cognitive research has often exhibited a negative association between calibration and age, with older adults showing tendencies of miscalibration, e.g. overconfidence (Castel, Middlebrooks, & McGillivray, 2016; Cauvin, Moulin, Souchay, Kliegel, & Schnitzspahn, 2019; Dodson, Bawa, & Krueger, 2007; Palmer, David, & Fleming, 2014; Soderstrom, Mccabe, & Rhodes, 2012). Cauvin and colleagues (2019) investigated age differences in calibration for an event-based prospective memory task (i.e. remembering to execute planned behaviour at the appropriate time in the future). Participants produced judgments of learning for both prospective (i.e. remembering there is something to do when a cue appears) and retrospective (i.e. remembering what do to do when the cue appears) components of the prospective memory task. There were no age differences in calibration for the retrospective component of the task. Conversely, old participants (average age of seventy years) were significantly overconfident in their performance judgments for the prospective component of the task, whereas young participants (average age of twenty-five years) were unbiased. Similarly, Palmer and colleagues (2014) found a negative relationship between age and metacognitive efficiency in a visual perceptual task, with confidence ratings made by older adults being less able to distinguish between correct and incorrect answers than ratings made by younger adults. Therefore, there is evidence to support the association of old age with poor calibration and overconfidence, possibly as a result of generalised age-induced cognitive decline that hinders older adults from making accurate performance judgments (Cauvin et al., 2019).

Nonetheless, age deficits in metacognition and calibration are not always present—even when deficits in cognition are (Castel et al., 2016; Cauvin et al., 2019; Hertzog & Dunlosky, 2011; Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Hertzog, Sinclair, & Dunlosky, 2010; Lin, Zabrucky, & Moore, 2002; Mcgillivray & Castel, 2011). For example, Hertzog and colleagues (2010) found that, despite recalling fewer items in a paired-associate recall test, older participants did not show lower metacognitive resolution (i.e. their judgments of learning could distinguish between information to be remembered or forgotten) compared to younger participants, as they gave lower judgments of learning that matched their performance. The discrepancy between studies that find age-related metacognitive decline and studies that do not is likely the result of numerous factors. Such factors can be differences in the cognitive and metacognitive mechanisms spared by age, metacognitive judgment format (e.g. using probabilities versus using intervals), strategic demands of a test, and task setting (Castel et al., 2016; Cauvin et al., 2019; Hansson, Rönnlund, Juslin, & Nilsson, 2008; Mcgillivray & Castel, 2011). Though all factors are important to consider, in the present review, I have focused on how task setting (i.e. whether the task is naturalistic or not) might serve as a mediator of the relationship between age and calibration, because of its potential implications for exercise research, where naturalistic settings are commonly implemented.

In laboratory and non-naturalistic examinations of prospective memory, older adults typically perform worse than younger adults (Henry, MacLeod, Phillips, & Crawford, 2004). However, in naturalistic settings and tasks, experience can play an important role in prospective memory performance through the implementation of metacognitive strategies that allow participants to complete the task successfully. Since older adults are assumed to have more experience in naturalistic prospective memory tasks (Cauvin et al., 2019), they would be expected to perform at least as well as younger adults. Interestingly, there is evidence to suggest that older adults exhibit better prospective memory in naturalistic settings than younger adults (Devolder, Brigham, & Pressley, 1990; Henry et al., 2004). In line with performance findings, we would also anticipate older adults to be better calibrated than younger adults in naturalistic tasks, resulting from higher experience with performance monitoring and evaluation. In accordance with this, Devolder and colleagues (1990) found that older adults made more accurate predictions about their performance in a naturalistic prospective memory task (i.e. making appointment calls at a pre-arranged day and time) compared to younger adults, who were overconfident. Their findings contrasted those of

Cauvin and colleagues (2019), who reported increased overconfidence in older adults compared to younger adults in a non-naturalistic laboratory prospective memory task. Therefore, though the effects of naturalistic versus non-naturalistic settings on the relationship between age and calibration need to be examined further, initial evidence suggests that task setting is a mediating factor. Non-naturalistic designs might not allow older participants to make use of their experience to optimise their performance and calibration, leading to apparent age deficits, which would not be observable in naturalistic designs.

In running, older runners are typically assumed to have more experience than younger runners (Knechtle, Rüst, Rosemann, Knechtle, & Lepers, 2012; Knechtle, Valeri, Zingg, Rosemann, & Rüst, 2014), especially since older adults without exercise experience might not always have the physical capacity to take up such a demanding activity. Because of this, we would expect older runners (and athletes in general) to be better calibrated than young runners—at least in naturalistic settings. In line with this prediction, Liverakos and colleagues (2018) found that older runners (>45 years old) were more precise and less underconfident in their half marathon finish time predictions than younger runners. However, the authors could not test whether experience markers such as years of running accounted for this relationship, as they had not collected relevant experience data. Overall, age has received limited attention in exercise calibration. To my knowledge, the only other studies that have examined the relationship between age and calibration in running have been inconclusive, either by failing to find evidence for the relationship or by observing a weak tendency for middle-aged runners to show less overconfidence than younger and older runners (Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016). Therefore, it is important to further explore how age and exercise calibration interact with each other. To do this, it would also be beneficial to include other experience factors to understand whether associations are driven by simple experience measures, such as years of training, or whether age can explain unique aspects of calibration. For example, older athletes might have to adjust their training in response to physical decline, leading to higher monitoring of and awareness regarding their abilities, and thus better calibration.

Research on the association between age and pacing in endurance sports has been more extensive than calibration research, and has generally shown that older runners are equally or more evenly paced than younger runners (Carlsson et al., 2016; Deaner, Addona, Carter, Joyner, & Hunter, 2016; Deaner et al., 2014; March et al., 2011; Nikolaidis, Cuk, Rosemann,

& Knechtle, 2019; Nikolaidis & Knechtle, 2017, 2018a; Trubee, Vanderburgh, Diestelkamp, & Jackson, 2014). Interestingly, when both age and experience associations with pacing were examined, the influence of age on reducing pace variation was present even after accounting for experience factors (Deaner et al., 2014). However, such research is scarce and insufficient in drawing reliable conclusions about the nature of the relationship between age and calibration. Nonetheless, pacing findings are useful in informing us about the influence of age on exercise metacognition and calibration, as they indicate that older runners are likely to adopt an even pacing profile in endurance sports, which likely reflects good exercise metacognitive skills and calibration.

1.3.1.4 Gender

Gender differences in bias have been explored in both cognitive and physical activities, and are thought to often generalise across domains (Gutierrez & Price, 2017). Cognitive studies have often shown patterns of male overconfidence and/or relative female underconfidence (Dahlbom, Jakobsson, Jakobsson, & Kotsadam, 2011; Gonida & Leondari, 2011; Gutierrez & Price, 2017; Jakobsson, 2012; Jakobsson, Levin, & Kotsadam, 2013; Lundeberg et al., 1994). Jakobsson (2012) asked university students to predict their grade for a macroeconomics test they were going to sit a week later. He observed that, though male students were unbiased in their predictions, female students were underconfident. Similarly, Dahlbom and colleagues (2011) asked 14-year-old school students to predict their performance on a mathematics test a week later, and found boys and girls to be overconfident and underconfident respectively. It should be noted that gender differences in bias could be influenced by various factors, which should be taken in consideration. These include type of task, social and cultural gender norms and stereotypes, as well as the extent to which a task is considered to be masculine, feminine, or gender neutral (Chiu & Klassen, 2010; Gutierrez & Price, 2017; Jakobsson et al., 2013; Lundeberg et al., 1994) . Nonetheless, there is substantial evidence to support the presence of male overconfidence and relative female underconfidence across a range of tasks and domains.

In exercise, gender differences in bias have been explored in running and physical education activities. Running calibration studies using point estimates for predictions (i.e. asking for a specific finish time) have generally shown that male runners are more overconfident than female runners in marathons and half marathons (Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018), supporting cognitive results. To my knowledge, Krawczyk and

Wilamowski (2018) have been the only authors to collect both marathon and half marathon data (using different participants). Interestingly, though male runners were more overconfident than female runners in both race distances, the differences in overconfidence between the two genders was smaller for half marathons compared to marathons. The researchers calculated bias score percentages relative to predicted finish time for both races, so this difference between the two distances was not simply the result of larger absolute differences in marathons relative to half marathons.

Conversely, two studies that examined gender differences in bias by collecting interval predictions (i.e. participants had to select a finish time interval out of available interval options) found women to be more overconfident than men in 10km and half marathon races (Liverakos et al., 2018; Nekby, Thoursie, & Vahtrik, 2008). However, neither study adjusted interval predictions for gender differences in performance, i.e. runners of both genders were presented with the same interval options instead of performance-adjusted intervals. Since women recorded slower finish times than men in both studies, their selected predictions were probably more likely to be in the upper (i.e. slower) range of each interval (Krawczyk & Wilamowski, 2018). Because of this, the likelihood of failing to achieve the predicted interval would be higher for women relative to men, leading to apparent overconfidence. Thus, findings of higher female relative to male overconfidence in running using interval predictions that have not been adjusted for performance differences are probably statistical artefacts. Overall, running studies that have employed point finish time predictions (i.e. participants estimated a specific finish time and did not just choose an interval) and controlled for gender differences in performance have found male runners to be more overconfident than female runners. Nonetheless, only a few studies have examined the impact of gender on running calibration, and thus further research is required to corroborate the relationship.

Research on gender differences in pacing has complemented running calibration results. Supporting patterns of relative male overconfidence, male runners tend to show more uneven pacing than female runners, by starting a race at a faster speed and slowing more later on (Cuk, Nikolaidis, & Knechtle, 2020; Deaner et al., 2016, 2014; Deaner & Lowen, 2016; March et al., 2011; Nikolaidis & Knechtle, 2018a; Smyth, 2018; Trubee et al., 2014). A potential explanation for this is that male runners are more likely to set overly optimistic goals, start a race too fast, and thus have to slow down more later on (Deaner et al., 2016;

Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018). In contrast, unbiased female runners are more likely to have realistic expectations about their performance, and adopt less risky pacing strategies. It should be noted, however, that most findings on gender differences in pacing have come from marathons, and might not necessarily generalise to other races.

Research on shorter races (e.g. 5km, 10km, and half marathons) has been relatively limited, and has shown that the tendency of male runners to slow more than female runners is less prominent than in marathons (Cuk et al., 2020; Deaner et al., 2016; Deaner & Lowen, 2016; Krawczyk & Wilamowski, 2018; Nikolaidis, Cuk, & Knechtle, 2019). Deaner and colleagues (2016) found that men were only slightly more likely to slow more in the second half of 10km races than women (2.0% versus 1.7% slowing). Additionally, Deaner and Lowen (2016) found that both genders paced a 5km race similarly when adjusting for performance based on finishing placement. Conversely, women were less likely than men to slow down when adjusting for performance based on finish time, though the magnitude of this difference depended on the degree of adjustment. More specifically, gender differences in pacing were present even after adjusting women's finish time by a theoretical value of 12%, but largely disappeared when gender differences in performance were fully accounted for by adjusting finish time by 21.5%.[2] This discrepancy in pacing between long and short races is in line with

[2] There are different approaches in accounting for gender differences in performance when examining gender effects on running bias and pace slowing. One approach is to match finish time intervals and conduct gender comparisons within each interval (e.g. Deaner et al., 2016, 2014; Deaner & Lowen, 2016). However, this approach requires a large sample size, which might not always be available. Another approach is to conduct multiple regressions, which include finish time as a factor, thus accounting for calibration or pacing variance associated with performance (e.g. Deaner et al., 2016, 2014; Krawczyk & Wilamowski, 2016). Deaner and colleagues (2014) argued that, for these approaches to be effective, women's finish times and finish time intervals need to be adjusted by a theoretical value of 12% to account for physiological differences in performance capacity between genders. When examining data from marathons, they found that, although men were more likely to slow during the race than women, adjusting women's finish times by 10%, 12%, and 16% led to an increasingly smaller difference in pace slowing between genders. These results indicated that, though gender differences in pace slowing are reliable, previous research has overestimated the magnitude of the effect.

Though this methodology aims to fully account for the effects of gender differences in performance capacity on bias and pace slowing, it is not without limitations. Pacing and bias studies using data from different race distances have often found inconsistent gender differences in performance. For example, women in the sample of 5km finishers analysed by Deaner and Lowen (2016) were slower than men by 21.5%; women in the sample of half marathon finishers analysed by Liverakos and colleagues (2018) were slower than men by 15%; and women in the sample of marathon finishers analysed by Deaner and colleagues (2014) were slower than men by 10%. Furthermore, gender differences in performance capacity might differ across exercise modalities, so it is not clear whether adjusting women's performance by 12% is appropriate for sports other than running. Given that in the thesis I examine gender differences in calibration across numerous running distances and high-intensity functional movement exercise (HIFME; see Chapter 3 for details) workouts, I have chosen to account for gender differences in performance using the multiple regression approach described above, but without adjusting women's performance. However, I acknowledge that this might lead to an overestimation of gender differences in calibration and that the development of appropriate, valid, and reliable methods to account for

Krawczyk and Wilamowski (2018) observation of lower male overconfidence in half marathons relative to marathons. This could indicate that relative male overconfidence is less pronounced in shorter races, where a fast starting speed might be easier to maintain throughout the course and where women do not benefit from slower glycogen depletion compared to men (Coyle, 2007; Deaner et al., 2016; Rapoport, 2010; Tarnopolsky, 2008). Nonetheless, gender differences in calibration and pacing have not been examined extensively enough across different race lengths. Thus, since different running distances have different physiological and pacing demands (del Coso et al., 2017; Gosztyla, Edwards, Quinn, & Kenefick, 2006; Nikolaidis, Cuk, & Knechtle, 2019; Nikolaidis, Cuk, Rosemann, et al., 2019; Smyth, 2018), it is important to further explore gender differences in bias across a range of races to better understand when and how they arise.

Physical education research has not provided evidence for gender contributions to calibration (Kolovelonis, 2019; Kolovelonis & Goudas, 2018; Kolovelonis, Goudas, & Dermitzaki, 2012), contrasting running and cognitive findings. Kolovelonis (2019) asked 210 male and 219 female school students, aged 10-12 years, to predict how many basketball shots they would make out of eight attempts during physical education classes. Overall, students of both genders overestimated how many shots they would make relative to their performance to a similar extent, indicating no gender differences in overconfidence. Similarly, Kolovelonis, Goudas, and Dermitzaki (2012) asked 40 male and 60 female 10-12-year-old students to predict how many cones they would dribble past with a basketball in 30 seconds. They found students of both genders to be similarly overconfident, with gender having no contribution to the degree of overconfidence. Overall, studies on physical education have not produced gender differences in bias. The reason behind the discrepancy between physical education and running findings is not clear, and might result from numerous factors, such as differences in age, task demands, and task type. Nonetheless, physical education findings are important in that they highlight the need to examine gender differences in bias across a range of exercise modalities. If gender contributes to bias in certain modalities but not others, then knowing which modalities these are will allow us to better understand when and how we can improve athlete calibration.

---

gender differences in performance capacity across exercise modalities and running distances is required to better assess the impact of gender on exercise calibration.

1.3.1.5 Summary of demographic factors

In Section 1.3.1, I reviewed evidence on the associations between demographic factors and calibration. Cognitive and exercise literature has suggested that expertise, experience, age, and gender all contribute to calibration, though each factor might do so in a different way and to a different extent. Understanding the exact nature and impact of their contributions is essential in assessing athlete calibration. Specifically, knowing how an athlete's expertise, exercise experience, age, and gender influence the accuracy of their performance monitoring and awareness can assist us in predicting whether the athlete will be well calibrated or not. Nonetheless, a common issue identified throughout the section was that the exercise literature on demographic factors and calibration has been relatively scarce and underdeveloped. Therefore, further research expanding on this literature is warranted before we are able to use demographic factors to make accurate and reliable assessments of athlete calibration.

1.3.2 Contributions of metacognition to calibration

Metacognition is a very important contributor to calibration. As discussed in Section 1.2.3.2, calibration is a measure of metacognition, as it assesses performance awareness accuracy. Changes in metacognitive behaviour (especially monitoring, reviewing, and evaluating performance) would thus be expected to lead to similar changes in calibration. For example, an athlete who is now monitoring their performance more than they used to, should become more accurate in their performance judgments. A limitation of the relationship between demographic factors and calibration reviewed in Section 1.3.1 is that demographic factors may be difficult and time-consuming (e.g. years of exercise experience), or even impossible (e.g. gender), to manipulate. Therefore, though we can use them to make assessments about an athlete's calibration, it is unlikely that we can devise interventions to alter them, and thus calibration, directly. Conversely, metacognition is malleable (Gutierrez & Schraw, 2015; Gutierrez de Blume, 2017; Nietfeld et al., 2006), rendering it feasible to implement metacognition manipulations that aim to improve calibration. In the present section, I review cognitive and, where available, exercise evidence on the relationship between metacognition and calibration. I first examine whether we can use offline self-reports of metacognition to inform us about calibration. Then, I review experimental studies that have measured the effects of metacognition manipulations on calibration directly.

1.3.2.1 Self-reports of metacognition and calibration

The strengths and limitations of offline measures of metacognition, such as self-reports, were discussed more extensively in Section 1.2.3.2. Despite validity issues (i.e. self-report questionnaires indicating metacognitive knowledge and regulation that might not correspond to behaviour; Veenman, 2011), an advantage of offline measures of metacognition is that they are highly cost-effective and time-efficient (Harrison & Vallin, 2018; Schellings & Hout-Wolters, 2011; Veenman, 2011; Winne & Perry, 2000). Because of this, a potential relationship between metacognition self-reports and calibration would allow us to use the former to make inferences about the latter in a quick and time-efficient manner. Coaches would then be able to utilise athlete scores on metacognition questionnaires to make judgments about their calibration.

Cognitive research has produced equivocal results on the presence of a relationship between self-reports of metacognition and calibration (Gutierrez & Schraw, 2015; Jacobse & Harskamp, 2012; Jang, Lee, Kim, & Min, 2020; Saraç & Karakelle, 2012; Schraw, 1997; Schraw & Dennison, 1994; Sperling, Howard, Staley, & DuBois, 2004; Tobias et al., 1999; Zepeda et al., 2015). Schraw and Dennison (1994) found that scores on the Metacognitive Awareness Inventory (MAI), which assesses metacognitive knowledge and regulation of cognition, did not correlate with calibration in a reading comprehension test in one hundred and ten undergraduate students. Similarly, Saraç and Karakelle (2012) found no correlation between Junior MAI scores and text comprehension calibration in forty-seven 9-11-year-old children, further supporting a dissociation between calibration and self-reports of metacognition.

In contrast, Schraw (1997) found that participants who scored low on the General Monitoring Strategies Checklist, which assesses self-reports of monitoring knowledge and strategies, were underconfident in lexical comparison, reading comprehension, syllogistic reasoning, and mathematics tests, whereas participants with moderate and high scores were unbiased. Tobias and colleagues (1999) examined the relationship of mathematics and verbal analogies knowledge-monitoring assessments, which measure the extent to which students are aware of what they do and do not know, with Learning and Study Strategies Inventory (LASSI; measures learning strategies) and MAI scores, and found weak positive correlations. Jang and colleagues (2020) observed a weak association between MAI scores and JOL (Judgment of Learning) absolute accuracy in a cued word recall task, as participants with higher MAI scores were more likely to produce accurate JOLs. The association between MAI scores and JOL absolute accuracy strengthened when participants completed the task a second and a third time. Interestingly, absolute accuracy scores in the first two task repetitions completely mediated this relationship for the third task repetition. This could mean that participants with high metacognitive scores are better able to gain task-specific experience through practice, which in turn leads to higher precision in future metacognitive judgments.

Overall, results on the relationship between self-reports of metacognition and calibration indicate the presence of a weak association. It appears likely that using questionnaires such as the MAI provides us with limited information on whether an individual is well-calibrated or not. It is not clear why there is a relative dissociation between the two types of measures, but it is likely the result of numerous factors. For example, online and offline measures of metacognition could capture different aspects of metacognition, or it could just be that individuals are able to exaggerate their metacognitive behaviour in self-reports, but not in calibration tasks.

The relationship between self-reports of metacognition and calibration in exercise settings has not been examined directly before. To my knowledge, only Nietfeld (2003) has explored whether metacognition self-reports correlated with pace monitoring, but not calibration, in physical activity (middle-distance running, which is an intriguing combination of aerobic and anaerobic exercise). Forty-five competitive middle-distance runners were provided with goal finish times at 80% of their fastest mile times, and had to run a mile as close to these times as possible. Participants also completed a Racing the Mile Questionnaire (RMQ), which was based on the MAI and assessed metacognitive strategy use and focus implemented while running a mile. Nietfeld found that participants who scored higher RMQ scores were more likely to finish close to their goal times, exhibiting heightened monitoring and regulating ability while running. Contrasting cognitive findings, these results indicated a moderate positive association between metacognitive strategy knowledge and monitoring accuracy in middle-distance running, suggesting that self-reports of metacognition might be useful in predicting exercise calibration.

The general discrepancy between cognitive results and Nietfeld's (2003) study could be partly explained by Nietfeld using a metacognition questionnaire whose items targeted the task examined directly. This is often not the case with metacognition self-reports in cognitive research, which typically use general items instead, possibly contributing to the observed lack of association with calibration (Schellings, 2011; Schellings, Hout-Wolters, Veenman, & Meijer, 2013). Nonetheless, it is not clear how specific self-report questionnaires would have to be for a potential relationship to become clearly visible (Jacobse & Harskamp, 2012). In any case, before we can infer that we can use metacognition self-reports to predict calibration in exercise, we need to collect further evidence. In Nietfeld's study, calibration was not examined directly, as participants did not generate their own performance predictions. Furthermore, runners were instructed to run at 80% of their best finish time, which did not equate to maximum effort, potentially rendering it easier for participants to maintain their pace throughout the trial. Studies replicating Nietfeld's findings whilst addressing these limitations are warranted before any strong conclusions can be reached.

1.3.2.2 The effects of metacognition manipulations on calibration

Our capacity to use self-reports of metacognition to assess athlete calibration might be fairly limited, but metacognition can still play a very important role in calibration. Cognitive research has shown that interventions aiming to improve metacognition have a positive effect on calibration (Griffin, Wiley, & Thiede, 2008; Gutierrez & Schraw, 2015; Gutierrez de Blume, 2017; Huff & Nietfeld, 2009; Nietfeld et al., 2006; Ramdass & Zimmerman, 2008; Thiede et al., 2003; Zepeda et al., 2015; Zimmerman et al., 2011). Nietfeld and colleagues (2006) found that undergraduate students became better calibrated following 16 weeks of monitoring training and calibration feedback during a college course. In contrast, participants who only generated self-feedback on their calibration without monitoring training did not experience the same benefits. The effect of metacognitive training on calibration was even more pronounced in the study conducted by Gutierrez and Schraw (2015), where just one hour of strategic training with an emphasis on metacognitive processes, such as monitoring and reviewing information, facilitated calibration and performance in a reading comprehension task compared to baseline. Calibration did not improve for participants in the control group who did not engage in metacognitive training. These results highlight the important role of metacognition in facilitating calibration, suggesting that even minimal interventions can improve calibration in cognitive and academic tasks. It should be noted, however, that to improve calibration interventions need to target metacognitive processes directly. Manipulations where individuals are merely provided with performance feedback and no information regarding their calibration might not be sufficient to elicit the desired outcomes (Hacker et al., 2012; Kolovelonis et al., 2013; Stone, 2000).

Research on the effects of metacognitive interventions on calibration in exercise has been limited, and consists of examinations of the effects of self-regulation training on calibration in physical education tasks in 10-12-year-old children (Kolovelonis, Goudas, & Dermitzaki, 2012; Kolovelonis et al., 2013, 2020). In two studies from the Kolovelonis lab (Kolovelonis, Goudas, & Dermitzaki, 2012; Kolovelonis et al., 2013), the researchers examined whether practicing a basketball-dribbling task under different self-regulatory conditions, i.e. where participants received performance feedback, set performance or process goals, and self-recorded their performance, would affect calibration in a basketball-dribbling test. Contrary to expectations of self-regulation training improving calibration, calibration in the self-regulation groups remained largely unaffected and was similar to that of control groups, which did not receive extensive self-regulation training. The authors argued that the lack of significant findings was likely the result of the interventions not targeting calibration directly.

Kolovelonis and colleagues (2020) expanded on the aforementioned studies by implementing a self-regulation intervention that targeted metacognitive processes linked to calibration during basketball-shooting practice. They asked participants in the experimental group to set their own practice goals, self-record these goals and their performance, engage in self-talk, self-reflect and self-evaluate their own performance, and make attributions about it. Conversely, the control group merely practiced the shooting task for the same sessions, and was not instructed to engage in the above self-regulatory behaviour. Results showed that the metacognitive manipulation was effective in improving calibration, as participants in the experimental group were more precise in their performance predictions following the intervention compared to the baseline test. Calibration did not improve for participants in the control group. Overall, findings from calibration research in physical education suggest that to improve calibration, it is important to devise interventions that target metacognitive processes related to calibration. Nonetheless, there is only a small number of studies on this relationship, which have only recruited young children. Developmental changes in metacognition and self-regulation render the extent to which we can generalise their results to populations of different ages unclear (Elferink-Gemser & Hettinga, 2017; Kolovelonis & Goudas, 2013; Kolovelonis, Goudas, & Dermitzaki, 2012; Wiersma, Stoter, Visscher, Hettinga, & Elferink-Gemser, 2017). Furthermore, as indicated by both exercise and cognitive research, interventions on metacognition can take numerous forms. To better understand how we can use metacognition manipulations to improve calibration, we need to conduct research that examines a wide range of interventions across ages. Doing so will have strong practical implications in allowing coaches to tackle athlete miscalibration effectively.

1.3.2.3 Summary of metacognition and calibration

In Section 1.3.2, I reviewed research on two aspects of the relationship between metacognition and calibration. First, I examined whether self-reports of metacognition can be used to inform us about athlete calibration. Cognitive research suggests that this is not always possible, but the one study that has examined this relationship in exercise observed a positive association between metacognition self-reports and pacing monitoring. Given the potential benefits of such a relationship, further research is required to address whether this finding is robust and also applies to calibration or not. Additionally, I reviewed research on the effects of metacognition manipulations on calibration. Both cognitive and exercise research support the use of such interventions to facilitate calibration. However, the absence of significant findings in some exercise studies also suggests that these interventions need to focus on calibration directly. Since not all metacognition and self-regulation manipulations are effective in improving calibration, more exercise research is required to test the effectiveness of different types of interventions.

1.4. CONCLUSIONS AND FUTURE DIRECTIONS

In Chapter 1, I explored two main topics. The first was the interaction between cognition and exercise, where I emphasized the impact of calibration, as a measure of metacognition, on physical activity. The second and main topic of the chapter was the examination of the influence of demographic factors and metacognition on calibration.

1.4.1 Conclusions from Part A

In Part A of the review, I discussed the importance of investigating calibration in exercise. Research on the relationship between sports and cognition has often focused on the beneficial effects of the former on the latter. However, there is an increasing number of studies to exhibit that this relationship is bidirectional, and that cognition plays an important role in exercise (e.g. cognitive abilities and fatigue can affect athletic performance). The role of calibration as a contributor to exercise has only recently started to receive attention, but has important implications. Based on previous cognitive and exercise findings and suggestions, I propose a series of ways in which calibration impacts exercise performance, motivation, and injury risk. More specifically, miscalibration can lead to the implementation of strategies that contribute to suboptimal performance. Suboptimal outcomes are also likely associated with low motivation for further engagement with an activity, whilst underconfidence can limit volition to attempt new and challenging tasks. At the same time, overconfidence and underconfidence can contribute to heightened injury risk by either increasing exposure to incident risk—this only applies to overconfidence—or by failing to induce adaptations resulting from appropriate training that minimise injury susceptibility. Therefore, since performance, motivation, and injury risk are essential components of exercise, Part A was able to highlight the necessity of identifying factors that contribute to calibration, and exploring ways in which we can facilitate it.

1.4.2 Conclusions from Part B

In Part B of the review, I explored two broad categories of constructs that can show associations with exercise calibration and thus warrant investigation. These were demographic factors (i.e. expertise, experience, age, and gender), and metacognition (self-reports and interventions). The literature showed that high expertise is associated with good prediction precision and minor underconfidence, whereas low expertise is associated with poor prediction precision and high overconfidence. Though cognitive and pacing research has consistently shown a positive influence of expertise on calibration, this has been less clear in exercise calibration research. It is thus important to conduct more studies that will allow us to examine and better understand whether expertise shows a similar influence on calibration across exercise modalities in the same way it does in cognition. Experience also showed a positive relationship with calibration across cognitive and exercise domains, with more experienced individuals being better calibrated than less experienced individuals. However, the reviewed literature suggested that individual experience markers only have a small contribution to calibration. This is likely because experience is a multifaceted factor that consists of numerous variables, meaning that each variable might only capture a small aspect of a broader experience construct. It is thus important to explore the influence of as many experience markers on calibration as possible to produce a comprehensive account of experience's influence on it.

Age was another contributor to calibration, though cognitive and exercise findings were not always in line. Cognitive studies in laboratories generally indicated that older adults were worse than, or just as well calibrated as, younger adults. However, there were suggestions that older adults might be able to use their experience to facilitate their calibration in naturalistic settings. Interestingly, exercise calibration and pacing studies indicated that older adults were often better calibrated and paced their races more evenly than younger adults, illustrating a positive contribution of age. Nonetheless, it is not clear whether this relationship could simply be the result of older athletes having more years of experience, with age influence diminishing after accounting for other experience factors. This needs to be tested directly to deduce whether the association between age and calibration is mediated by other experience factors, or whether age has an independent relationship with calibration—possibly as a separate experience factor.

One of the most well examined factors in cognitive and exercise calibration was gender. Overall, there is strong evidence from cognitive and marathon running studies to suggest that men are more overconfident and less underconfident than women. However, a limited number of calibration and pacing studies in running has suggested that this effect might be less pronounced in shorter race distances, where the gender differences in bias were smaller compared to marathons. This could be attributed to physiological differences between the two genders, as well as different demands in races of different lengths. It is thus important to examine the reliability of these gender differences across race distances. Furthermore, physical education findings have not exhibited gender differences in bias, suggesting that gender influence on calibration needs to also be examined in other exercise modalities.

The second category of constructs I reviewed was related to metacognition, and referred to metacognition self-reports and interventions. Since metacognition and calibration are closely related, we would expect patterns and changes in the latter to reflect patterns and changes in the former. I thus examined literature on whether self-reports of metacognition could inform us about an athlete's calibration. This would have important implications for coaches, as it would allow them to use self-report questionnaires to assess their athlete's performance awareness in a time-efficient manner. Cognitive research suggested that such a relationship is weak at best, as most studies have not observed an association between metacognition self-reports and calibration. Interestingly, the one exercise study that explored this relationship indirectly, using running self-reports and pacing monitoring, indicated a moderate correlation between the two. However, to my knowledge, there has been no study to examine this relationship in exercise using calibration. Therefore, more research is required to establish whether metacognition self-reports can actually inform us regarding athlete calibration in exercise.

I also reviewed research on the extent to which we can use interventions targeting metacognition to facilitate calibration. Cognitive research demonstrated that experimental interventions of metacognition can lead to improvements in calibration. However, research on this relationship is very limited in exercise, and only one out of the three studies available has illustrated a positive effect of self-regulation training targeting metacognitive processes on calibration. This is likely the result of the two studies that did not report significant findings not implementing interventions that targeted calibration directly. The implications of these results are intriguing, because they suggest that we can use manipulations of metacognitive processes (e.g. monitoring, reviewing, and evaluating performance) to facilitate calibration. Since this relationship has received very little attention in exercise, we need more studies to examine the different types of interventions that have the capacity to improve calibration.

1.4.3 Thesis content and structure

Overall, Chapter 1 explored whether and how demographic factors and metacognition are associated with calibration. It also identified areas of research that require further examination through correlational and experimental studies. The present thesis aims to address these areas to expand our understanding of the variables that contribute to exercise calibration and address the limitations of previous work.

More specifically, Chapters 2 and 3 contain correlational examinations of the associations between demographic factors (i.e. expertise, experience, age, and gender) and calibration in running and high-intensity functional movement exercise (HIFME; e.g. CrossFit and circuit training). In Chapter 2, I expand on previous work by collecting demographic and prediction data from 10km and half marathon races, whilst controlling for prediction factors (e.g. prediction type, and time when predictions were made before each race). Based on the literature reviewed in Section 1.3.1, I anticipate that runners with higher expertise and experience, and older age will be better calibrated than runners with lower expertise and experience, and younger age. I am also interested in the extent to which male runners will be more overconfident than female runners in line with marathon findings.

In Chapter 3, I explore associations between demographic factors and calibration in HIFME, because HIFME consists of more complex activity patterns than running, meaning that performance might be more difficult to predict. Similar to Chapter 2, I expect expert and experienced athletes to be better calibrated than non-expert and less experienced athletes. Given the lack of gender differences in physical education calibration studies, I am uncertain as to whether gender differences in bias will arise in HIFME workouts.

In Chapters 4 and 5, I focus on the relationship between metacognition and calibration in HIFME and running. In Chapter 4, I examine whether we can use self-reports of metacognition and cognitive calibration to predict HIFME and running calibration. I anticipate that task-specific questionnaires will likely be better predictors of exercise calibration than general exercise questionnaires, as there are suggestions that task-specificity is important for a relationship between offline and online measures of metacognition to arise (Schellings, 2011; Schellings et al., 2013). I also expect that domain-general metacognitive components will lead to positive correlations in calibration between cognitive and exercise domains (I present relevant research in Sections 4.1 and 4.4.1).

In Chapter 5, I present and analyse data from two studies, where I experimentally manipulate the instructions I provide to participants on how to make their performance predictions. These studies test whether strategic approaches to making predictions lead to better HIFME and running calibration than impulsive approaches. In the study examining HIFME calibration (see Study 7), I provide detailed instructions for strategic and impulsive predictions to examine whether such a manipulation can improve HIFME calibration in participants without HIFME experience. In the running calibration study (see Study 8), I only provide participants with minimal and non-specific instructions for strategic and impulsive predictions to examine whether, despite the lack of specific instructions, athletes with previous running experience will still engage in effective strategic thinking, and thus exhibit better calibration compared to impulsive predictions. I expect that athletes in both studies will be better calibrated when asked to make strategic, rather than impulsive, predictions.

I discuss the implications and limitations of the research presented in the next four chapters in Chapter 6, which is the final chapter of the thesis.

CHAPTER 2: DEMOGRAPHIC FACTORS & RUNNING CALIBRATION

2.1 INTRODUCTION & RATIONALE

In Section 1.2.3, I discussed the importance of calibration in exercise performance, motivation, and injury risk. In doing so, I highlighted the need to examine variables that contribute to good and poor calibration. Demographic factors in exercise refer to athlete characteristics, and have the potential to inform us about calibration in a range of physical activities. By reviewing cognitive and exercise calibration, as well as pacing, research in Section 1.3.1, I identified expertise, experience, age, and gender as demographic factors that can be associated with exercise calibration. Using this information, athletes and coaches alike can carry out initial calibration assessments, enabling them to adjust predictions and strategies appropriately, in turn leading to performance optimisation. However, the reviewed empirical work also demonstrated the need to conduct more research on the relationship between these demographic factors and exercise calibration before we can use the former in such a manner.

The finding that higher expertise in associated with better prediction precision and lower overconfidence is robust in cognitive research, and has also gathered considerable support from pacing studies (see Section 1.3.1.1). In the limited running calibration research available, there have been suggestions that faster running speed is associated with lower overconfidence in races. However, this evidence is weak and inconsistent, possibly because of methodological limitations. For example, runners who did not finish a marathon in less than 6 hours were not part of the sample analysed by Hubble and Zhao (2016), and, since they were most likely to be in the slowest race corral (which indicated low expertise), the bias of the corral group could have been underestimated. Furthermore, to my knowledge, there has been no examination of the relationship between expertise and running prediction precision, with running research focusing on bias instead (e.g. Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016). Because of this, it is important to conduct analyses that examine the extent to which faster finish times are associated with better running calibration. Assuming a link between expertise and experience, it would also be interesting to include experience factors in expertise analyses to better understand the individual calibration contributions of each factor.

There is considerable evidence to show a positive influence of experience on calibration in cognition and exercise (see Section 1.3.1.2). Experience is a multifaceted construct that consists of numerous factors. Thus, to be able to fully understand its impact on exercise calibration, we need to consider a wide range of experience markers—especially since individual markers often only exhibit small contributions to calibration. So far, there is limited evidence on the positive influence of experience (through club membership and race repetition) on running calibration. Since research on cognition, pacing, and other exercise modalities (e.g. basketball) has exhibited a role for more experience variables, such as years of experience and training volume, their influence on running calibration should also be examined to test their generalisability across domains and exercise modalities.

Age has the potential to serve as an experience marker (see Section 1.3.1.3). Though older adults can be overconfident and less precise in their performance judgments than younger adults in controlled laboratory studies, higher experience with age can produce the opposite results in naturalistic settings. In line with this, running calibration and pacing studies have produced some evidence that older runners are better calibrated and more (or at least as) evenly paced than younger runners. However, calibration and pacing studies have very rarely accounted for the effects of other experience factors when assessing age associations with calibration and pacing, leaving open the possibility of age findings resulting from other experience factors. For example, older runners might have just been running for more years than younger runners. In that case, the inclusion of years of running experience in the analysis should eliminate age influence on calibration. Is it thus important to explore the relationship between age and running calibration while accounting for other experience markers to deduce whether age can serve as an independent experience factor or whether its influence on calibration is simply the result of other experience factors.

Gender has consistently been shown to contribute to calibration bias—and pacing—in cognition and running, with men showing patterns of overconfidence relative to women (see Section 1.3.1.4). Interestingly, a small number of running calibration and pacing studies seem to suggest that gender differences are less pronounced in races shorter than marathons. Though there is no definitive explanation for this pattern of results, it is likely attributable to differences in physiological and psychological demands associated with different race lengths. Since the majority of running studies on calibration and pacing have focused on marathons, further examinations on the presence and magnitude of gender differences in bias should be conducted across a range of race lengths, e.g. half marathons and 10ks. Results from such examinations should improve our understanding on male and female bias tendencies in running.

The aim of Chapter 2 was to address the gaps identified in the reviewed literature on the relationship between demographic factors and running calibration. To achieve this, I analysed demographic and calibration data collected from recreational runners participating in the 2018 Edinburgh Christmas 10km Run and the 2018 Alloa Half Marathon. In both studies, I asked participants to provide demographic information regarding their previous running experience (i.e. training volume, years of running experience, and club membership), age, and gender. I determined expertise based on race finish time (for a review on issues with this operationalisation of expertise, see Section 1.3.1.1). In contrast to previous running calibration studies, where participants were only asked to indicate their expected finish times (Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018; Liverakos et al., 2018), in the present two studies, participants also provided their goal predictions. I did this to control for the possibility of runners giving goal predictions when asked to make predictions about their expected finish times (hereon referred to as "realistic predictions" to contrast "goal predictions", though it should be noted that goal predictions can also be realistic). Additionally, I was interested in examining potential dissociations between the two prediction types, which could help indicate whether runners benefit more from making realistic or goal predictions.

Across studies, I anticipated that faster, more experienced (i.e. club members, runners with more years/months of running experience, and runners who ran more kilometres per week), and older runners would be better calibrated than slower, less experienced, and younger runners. Collecting numerous and novel markers of experience afforded the opportunity to examine sources of unique variance and expand on previous experience and age research, allowing me to determine whether other markers of experience drive positive age influence in calibration and pacing. To test the generalisability of relative male overconfidence observed in marathon predictions and pacing to shorter races (10km in Study 1 and half marathon in Study 2), I examined gender differences in bias, possibly anticipating small or non-significant associations. Finally, in my analysis of prediction type effects on calibration, I expected runners to demonstrate a tendency towards overconfidence in their goal predictions (Sackett, Wu, White, & Markle, 2015), and to be less overconfident in their realistic predictions.

2.2 STUDY 1 – EDINBURGH CHRISTMAS 10K RUN

2.2.1 Study specifics

In Study 1, I collected data from the 2018 Edinburgh Christmas 10k Run, a timed 10km race at near sea level, with a total ascent of approximately 50m, and 442 finishers. Participants were recreational runners of various expertise levels. The race took place on the 2nd of December 2018, and the temperature in Edinburgh at the time ranged from 5° to 6°C. To my knowledge, this was the first study to examine the association between demographic factors and calibration in a 10km race using point predictions (i.e. runners predicted a specific finish time). Only Nekby and colleagues (2008) have explored calibration in a 10km race before, but they collected interval predictions (i.e. participants were given the option to select predetermined finish time intervals), and only focused on gender differences. Given that races of different lengths have different physiological and pacing demands (del Coso et al., 2017; Gosztyla et al., 2006; Nikolaidis, Cuk, & Knechtle, 2019; Nikolaidis, Cuk, Rosemann, et al., 2019; Smyth, 2018), it was important to examine how demographic factors influence running calibration across a range of distances. Therefore, Study 1 aimed to further our understanding of running calibration by increasing the number of demographic factors explored in a popular, but scarcely examined, race distance. My predictions for Study 1 results were consistent with those described at the end of Section 2.1.

2.2.2 Methods

2.2.2.1 Participants

Two-hundred-and-seven runners gave their finish time predictions for the 10km race in the 24 hours preceding it. I removed data from six runners who either failed to finish the race, or their predictions could not be matched to their finish times (e.g. their name was missing on the prediction form). Two runners failed to provide realistic predictions and one runner failed to provide a goal prediction, so I excluded them from all analyses involving realistic and goal predictions respectively. I also removed data from seven (six women and one man; seven unaffiliated runners) and four (two women and two men; four unaffiliated runners) outliers from the realistic and goal prediction analyses respectively. I defined outliers according to realistic and goal prediction absolute accuracy percentage scores. I considered participants to be outliers, if the absolute value of their absolute accuracy *z*-scores exceeded three. I removed nine outliers (six women and three men; nine unaffiliated runners) from the within subjects (prediction type) analyses, and two outliers (two women; two unaffiliated runners) from finish time analyses.

Overall, I analysed data from 199 runners (95 male and 104 female runners; 38 club and 161 unaffiliated runners; $M = 38.1$ years, $SD = 10.8$ years), contributing 192 realistic predictions (94 male and 98 female runners; 37 club and 155 unaffiliated runners; $M = 38.2$ years, $SD = 10.8$ years) and 196 goal predictions (92 male and 104 female runners; 38 club and 158 unaffiliated runners; $M = 38.0$ years, $SD = 10.7$ years). I compared prediction types using data from the 189 participants that provided a full set of data without outliers (91 male and 98 female runners; 37 club and 152 unaffiliated runners; $M = 38.1$ years, $SD = 10.7$ years).

The present study secured ethical approval from the University of St Andrews School of Psychology & Neuroscience Ethics Committee (Ethics approval code: PS13950; see Appendix 8.1.1) and was in accordance with the Declaration of Helsinki. Participants did not receive compensation for their participation.

2.2.2.2 Materials

I collected the following identifying and demographic information using a short questionnaire (see Appendix 8.2): name; date of birth; club membership; kilometres run per week; and years and months of running experience. On the same questionnaire, participants gave two 10k finish time predictions. The first prediction was their goal time, and instructions were: "The finish time I hope to achieve (my goal time) is:". The second prediction was their realistic time, and instructions were: "The finish time I think is most likely for me to achieve is:". The questionnaire instructed participants to read instructions for both prediction types before making their finish time estimates to control for order effects. I collected gender and finish time data online from the results published after the race, which I matched to the questionnaire data using the identifiers provided by the participants.

2.2.2.3 Design

The focus of the analysis was on the influence of expertise, experience, age, and gender on calibration. I determined calibration by calculating bias (subtracting actual finish time from predicted finish time) to assess direction and magnitude of over/underconfidence, and absolute accuracy (absolute value of bias) to assess precision. I assessed expertise based on finish time, and quantified experience in the three following ways: kilometres run per week; months of running experience; and club membership. Participants made predictions within a window of 24 hours prior to the performance measure to control for confounds linked to the time before the race when predictions were made. Twenty-nine participants gave their realistic and goal predictions the day before the race, and 163 and 167 gave their realistic and goal predictions respectively on the day of the race (29 and 160 respectively for analyses comparing prediction type).

In calibration analyses, I accounted for the effects of finish time value variation on calibration by creating bias and absolute accuracy score percentages based on each participant's finish time (i.e. (*bias/finish time*) × 100 and (*absolute accuracy/finish time*) × 100). This allowed me to interpret bias and absolute accuracy scaled by performance time on the basis that a 5-minute prediction discrepancy for someone running the 10k in 35 minutes represents a greater metacognitive error than a 5-minute discrepancy for someone running the 10k in 70 minutes.

2.2.2.4 Procedure

To administer questionnaires, a team of University of St Andrews researchers approached athletes when they collected their running chips prior to the race. Runners who took part in the study provided their verbal informed consent before completing the study questionnaire (took approximately 1-2 minutes). There was very limited information available on the runners who did not wish to participate. After completing the questionnaire, the research team gave participants information on where to find debrief information online. I used identifying information to match the prediction data to the finish times published online following the conclusion of the race.

2.2.3 Results

2.2.3.1 Data checks

I first compared the finish times in the 2017, 2018, and 2019 Edinburgh Christmas 10k Runs to test whether the mean performance in the 2018 race examined in the present study was similar with performance in the same course in other years. I then compared finish times between runners who participated in the study and runners who did not to test whether participants had a different profile to non-participants. I also examined prediction type frequencies to test whether participants were more likely to make faster goal or realistic predictions.

Differences in finish time between the 2017, 2018, and 2019 Edinburgh Christmas 10k Runs. The mean finish time of the 2018 race examined in Study 1 was 54 minutes and 44 seconds (*SD* = 8 min 46 s). The mean finish time of the 2017 race was 56 minutes and 46 seconds (*SD* = 9 min 47 s), and the mean finish time of the 2019 race was 57 minutes and 31 seconds (*SD* = 9 min 52 s). A one-way between subjects ANOVA showed a significant difference between the finish times from the three years, $F_{(2, 1391)} = 10.64$, $p < .001$, $\eta p^2 = .015$. Pairwise comparisons using a Sidak correction demonstrated that the 2018 race was significantly faster than both the 2017 ($p = .004$) and the 2019 ($p < .001$) races. There was no significant difference between 2017 and 2019, $p = .530$. These results indicate that other factors, e.g. weather conditions, could have contributed to the 2018 race examined here being faster than the 2017 and 2019 races (there were no data available for other years, e.g. 2016). Runners who relied on previous course experience to make their predictions could have thus made more underconfident and less overconfident estimates compared to their actual performance in anticipation of harsher racing conditions, with potential implications for calibration findings.

Differences in finish time between sample and non-participants. The mean finish time of the 241 runners who finished the 10km race but did not participate in Study 1 was 56 minutes and 1 second (*SD* = 8 min 55 s). The average finish time of the 199 runners comprising the sample (~45% of the field of 442 runners) was 53 minutes and 2 seconds (*SD* = 8 min 14 s), indicating that participants in the sample were significantly faster than the runners who did not participate in the study $t_{(438)} = 3.61$, $p < .001$, $d = 0.35$.

Day before versus day of the race. There were no differences in finish time (day before: *M* = 52 min 24 s, *SD* = 9 min 26 s: day of the race: *M* = 53 min 8 s, *SD* = 8 min 2 s), realistic predictions (day before: *M* = 53 min 30 s, *SD* = 9 min 28 s; day of the race: *M* = 55 min 3 s; *SD* = 9 min 1 s), and goal predictions (day before: *M* = 52 min 3 s, *SD* = 8 min 51 s; day of the race: *M* = 53 min 18 s, *SD* = 8 min 12 s) between participants who gave their predictions on the day before the race and participants who gave their predictions on the day of the race, all *ps* > .10.

Prediction type frequencies. Of the 189 participants who provided valid realistic and goal predictions, only nine participants (~5%) made realistic predictions that were faster than their goal predictions; fifty-five (~29%) made the same realistic and goal predictions; and one-hundred and twenty-five (~66%) made goal predictions that were faster than their realistic predictions. A chi-squared test indicated that there was a significant difference in the number of participants who made goal predictions that were faster, equal, or slower compared to realistic predictions, $\chi^2_{(2)} = 108.32$, $p < .001$. Thus, runners in the sample were more likely to set finish time goals that were faster than the finish times they expected to achieve in the race.

2.2.3.2 Performance & predictions

To examine the associations of experience factors (kilometres run per week, months of running experience, and club membership), age, and gender with finish time, realistic predictions, and goal predictions, I conducted correlations and multiple regressions. In the latter, I entered all factors as predictors at the same time to account for shared variance.

Table 2.1

Descriptive statistics for Club membership.

	Club		**Unaffiliated**	
Outcome Variable	***Mean***	***SD***	***Mean***	***SD***
Finish Time	50 min 22 s	9 min 9 s	53 min 39 s	7 min 54 s
Realistic Prediction	51 min 59 s	10 min 25 s	55 min 29 s	8 min 37 s
Goal Prediction	51 min 22 s	10 min 22 s	53 min 32 s	7 min 41 s
Realistic Bias	3.48%	7.11%	3.99%	7.38%
Realistic Absolute Accuracy	5.83%	5.31%	6.32%	5.50%
Goal Bias	1.82%	6.03%	0.00%	6.68%
Goal Absolute Accuracy	4.37%	4.48%	5.09%	4.32%

Note. The table provides information on the means and standard deviations (SD) of club members and unaffiliated runners for performance, predictions, and calibration.

Table 2.2

Descriptive statistics for Gender.

	Male		Female	
Outcome Variable	***Mean***	***SD***	***Mean***	***SD***
Finish Time	48 min 30 s	7 min 4 s	57 min 9 s	6 min 59 s
Realistic Prediction	49 min 42 s	7 min 7 s	59 min 43 s	8 min 0 s
Goal Prediction	48 min 21 s	7 min 0 s	57 min 20 s	6 min 58 s
Realistic Bias	2.89%	8.16%	4.85%	6.28%
Realistic Absolute Accuracy	6.21%	6.01%	6.24%	4.89%
Goal Bias	0.12%	6.78%	0.57%	6.44%
Goal Absolute Accuracy	5.06%	4.49%	4.85%	4.24%

Note. The table provides information on the means and standard deviations (SD) of male and female runners for performance, predictions, and calibration.

Finish time. Results from correlational analyses and the multiple regression model for performance can be seen in Tables 2.3 and 2.4. The performance variance explained by the regression model was significant, $R^2 = .494$, $F_{(5, 189)} = 36.94$, $p < .001$. Runners who ran more kilometres per week and had been running for longer were faster to finish the race than less experienced runners. Though club members were faster to finish the race when the correlation of club membership with performance was examined (Table 2.1), club membership contributed little unique variance beyond the shared variance accounted for by training volume and months of running in the regression model. Conversely, after accounting for the other experience factors, age became a significant predictor of finish time, with older runners being slower finishers than younger runners, whereas this relationship was not significant when the correlation between age and finish time was examined. Male runners were faster to finish the race than female runners across analyses (Table 2.2).

Table 2.3

Correlation coefficients for the associations between demographic factors and the outcome variables finish time, realistic predictions, and goal predictions.

	Performance		Realistic		Goal	
Factor	*r*	*p*	*r*	*p*	*r*	*p*
km	-0.53	< .001	-0.50	< .001	-0.48	< .001
Months	-0.22	.002	-0.22	.003	-0.20	.006
Club	-0.16	.026	-0.15	.034	-0.10	.148
Age	0.01	.893	0.04	.605	0.10	.189
Gender	-0.53	< .001	-0.55	< .001	-0.54	< .001

Note. *r* represents the correlation coefficient of each factor with the outcome variables. *p* represents the *p*-value associated with corresponding predictor and outcome variable. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, and Gender to runner gender (Gender = 0, if female; Gender = 1, if male).

Table 2.4

Multiple regression coefficients for demographic predictors on the outcome variable finish time.

Coefficient	***B***	***Beta***	**Std. Error**	***t***	***p***
Intercept	3466.80	—	100.62	34.46	< .001
km	-14.51	-0.41	2.00	-7.26	< .001
Months	-0.77	-0.18	0.24	-3.21	.002
Club	-65.30	-0.05	69.91	-0.93	.351
Age	8.22	0.18	2.61	3.15	.002
Gender	-428.30	-0.43	53.52	-8.00	< .001

Note. Demographic predictors were entered at the same time in the multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, and Gender to runner gender (Gender = 0, if female; Gender = 1, if male).

Predictions. Results for correlational analyses and the multiple linear regressions models for both prediction types can be seen in Tables 2.3 and 2.5. The variance explained by each regression model was significant: realistic predictions—R^2 = .506, $F_{(5, 182)} = 37.22$, $p < .001$; goal predictions—R^2 = .485, $F_{(5, 186)} = 35.04$, $p < .001$. Mirroring finish times findings, demographic factors, with the exclusion of club membership, significantly contributed to both prediction types. Runners who ran more kilometres per week, had been running for longer, were younger, and male were more likely to make faster realistic and goal predictions than less experienced, older, and female runners (Table 2.2). As with performance findings, age was only a significant contributor to predictions after accounting for other experience factors, and not when examined individually. Club membership correlated with realistic predictions, though not with goal predictions (Table 2.1), and did not make a significant contribution to either regression model after accounting for shared variance. Overall, experience, age, and gender exhibited consistent contributions to performance, and realistic and goal predictions.

Table 2.5

Multiple regression coefficients for demographic predictors on the outcome variables realistic and goal predictions.

Prediction type	**Coefficient**	***B***	***Beta***	***Std. Error***	***t***	***p***
Realistic	Intercept	3575.99	—	112.14	31.89	< .001
	km	-14.67	-0.38	2.19	-6.70	< .001
	Months	-0.90	-0.19	0.26	-3.42	.001
	Club	-86.75	-0.06	76.69	-1.13	.259
	Age	10.05	0.20	2.88	3.48	.001
	Gender	-503.64	-0.46	58.94	-8.55	< .001
Goal	Intercept	3325.06	—	103.63	32.09	< .001
	km	-13.22	-0.37	2.04	-6.48	< .001
	Months	-0.79	-0.18	0.24	-3.22	.002
	Club	-27.82	-0.02	71.01	-0.39	.696
	Age	11.57	0.25	2.69	4.30	< .001
	Gender	-450.18	-0.45	54.81	-8.21	< .001

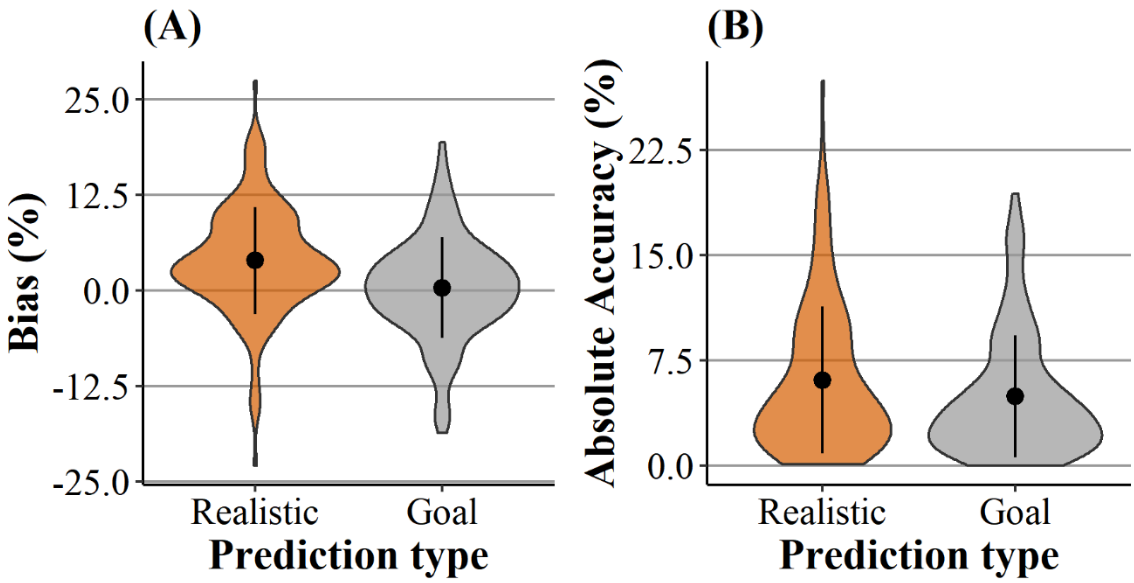
Note. Demographic predictors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, and Gender to runner gender (Gender = 0, if female; Gender = 1, if male).

2.2.3.3 Calibration – Prediction type

Bias. In Studies 1 and 2, bias values close to zero indicate low overconfidence or underconfidence, positive values indicate underconfidence, and negative values indicate overconfidence. To test differences in bias between goal and realistic predictions, I conducted a paired samples t-test, which showed them to be significantly different, $t_{(188)} = 8.89$, $p <$ .001, $d = 0.63$. Two one-sample t-tests comparing realistic and goal prediction bias to zero, i.e. no bias, showed that realistic predictions were significantly underconfident ($M = 3.92\%$, $SD = 7.02\%$), $t_{(188)} = 7.67$, $p < .001$, $d = 0.56$, whilst goal predictions were not significantly biased ($M = 0.35\%$, $SD = 6.59\%$), $t_{(188)} = 0.72$, $p = .471$, $d = 0.05$ (Fig. 2.1A).

Figure 2.1

Violin plots illustrating the effects of prediction type on bias and absolute accuracy.



Note. Panel A illustrates prediction type effects on bias. Panel B shows prediction type effects on absolute accuracy. The perimeter of each violin plot illustrates density, the central point represents the mean, and the vertical line represents +/- one standard deviation.

Absolute accuracy. Absolute accuracy values close to zero indicate a low prediction-performance discrepancy, and thus higher precision. To test differences in absolute accuracy between goal and realistic predictions, I conducted a paired samples t-test, which showed them to be significantly different, $t_{(188)} = 3.33$, $p = .001$, $d = 0.22$. Goal absolute accuracy ($M = 4.94\%$, $SD = 4.36\%$) was smaller than realistic absolute accuracy ($M = 6.08\%$, $SD = 5.24\%$), indicating higher precision for goal predictions (Fig. 2.1B). Overall, runners made realistic predictions that were underconfident and with poorer precision, and goal predictions that were unbiased and more precise.

2.2.3.4 Calibration – Demographic factors

2.2.3.4.1 Realistic Predictions

Bias. To examine the extent to which demographic factors can predict bias, I conducted correlations and two multiple regression models in which I entered all factors as predictors at the same time to account for shared variance. In the first regression model, the factors I entered were training volume, months of running experience, club membership, age, and gender. In the second model, I also included finish time as a measure of expertise. I did this because finish time is associated with other demographic factors (as seen in Section 2.2.3.2), and I wanted to investigate the influence of each demographic factor on bias with and without its presence in the model. I implemented the same method of analysis for both bias and absolute accuracy, and for both prediction types.

Table 2.6

Correlation coefficients for the associations between demographic factors and the outcome variables bias and absolute accuracy for realistic and goal predictions.

	Real Bias		**Real Abs Acc**		**Goal Bias**		**Goal Abs Acc**	
Factor	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***
km	-0.04	.618	-0.20	.006	0.10	.186	-0.20	.004
Months	-0.02	.771	-0.09	.202	0.06	.421	-0.06	.439
Club	-0.03	.706	-0.04	.621	0.11	.127	-0.07	.366
Age	0.03	.647	-0.14	.064	0.13	.069	-0.18	.014
Gender	-0.13	.064	0.00	.966	-0.03	.637	0.02	.743
FinTime	-0.11	.116	0.12	.112	-0.21	.003	0.21	.003

Note. *r* represents the correlation coefficient of each factor with the outcome variables. *p* represents the *p*-value associated with corresponding predictor and outcome variable. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), and FinTime to finish time.

Results from multiple regression models were not in line with correlations examining the individual association of each factor with realistic prediction bias (see Table 2.6 for correlation coefficients and Table 2.7 for the regression models). Though the multiple regression model that did not include finish time was not a significant predictor of bias, R^2 =

.016, $F_{(5, 182)} = 0.61$, $p = .695$, the model that included finish time was, $R^2 = .091$, $F_{(6, 181)} = 3.01$, $p = .008$. Faster and female runners were significantly more likely to be underconfident than slower and male runners respectively (Fig. 2.2A & 2.2C). When examining the distribution of male and female runner bias relative to finish time, Figure 2.2C shows that, at faster finish times, men and women showed similar underconfidence. However, with increasingly slower finish times, men became less underconfident, and even exhibited some slight overconfidence for the slowest finish times, whereas women were consistently underconfident, suggesting that gender differences in bias were driven by slower finish times. Training volume showed a non-significant tendency to predict bias in the model that included finish time, with higher training volume predicting lower underconfidence (Fig. 2.2B). Months of running experience, club membership, and age did not make significant contributions to either model.

Table 2.7

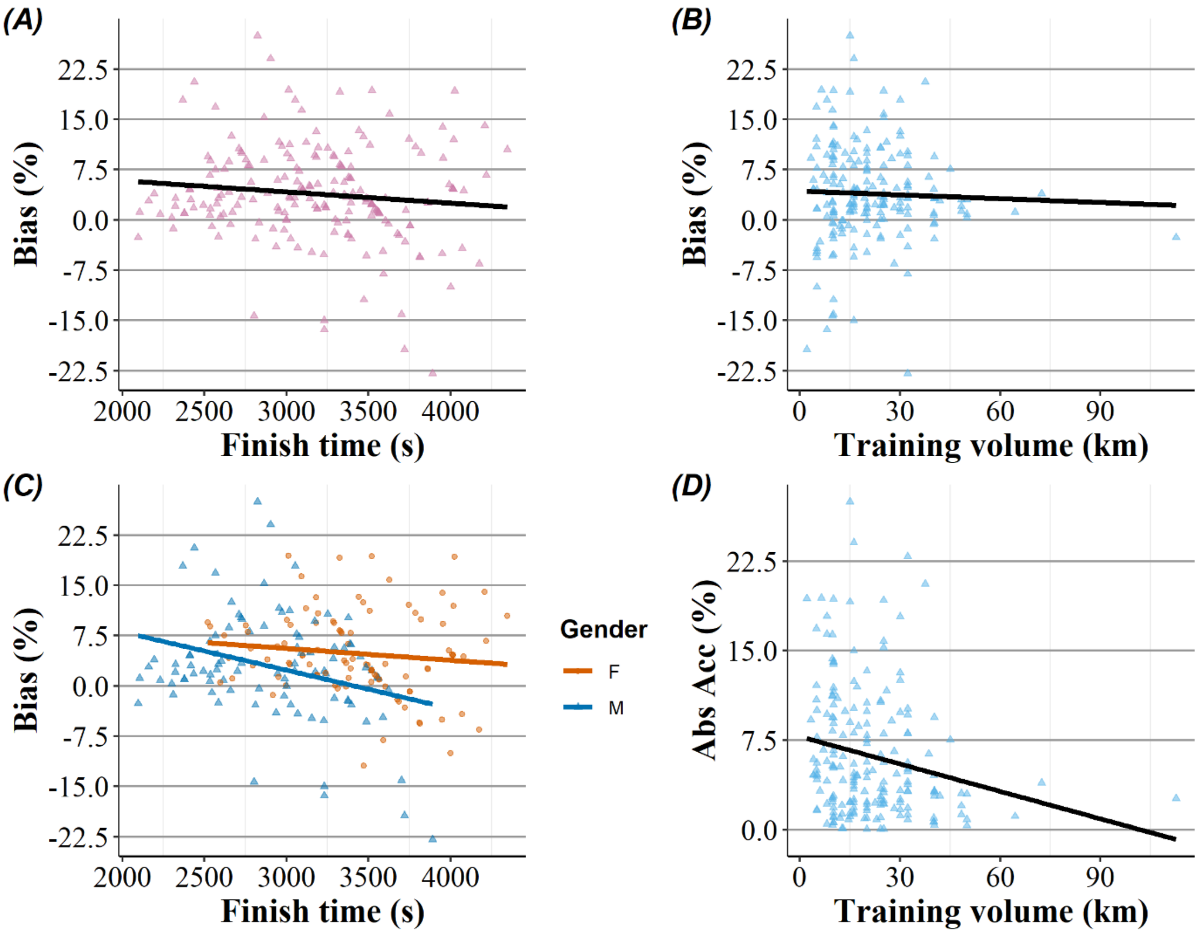
Multiple regression coefficients for demographic predictors on the outcome variable realistic prediction bias.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No finish time	Intercept	4.07	—	2.08	1.96	.052
	km	-0.01	-0.02	0.04	-0.19	.848
	Months	0.00	-0.03	0.01	-0.40	.689
	Club	-0.68	-0.04	1.42	-0.48	.634
	Age	0.03	0.05	0.05	0.60	.547
	Gender	-1.58	-0.11	1.09	-1.45	.150
Finish time included	Intercept	22.91	—	5.29	4.33	< .001
	km	-0.09	-0.17	0.04	-1.94	.053
	Months	-0.01	-0.10	0.01	-1.30	.195
	Club	-1.04	-0.06	1.37	-0.76	.451
	Age	0.08	0.12	0.05	1.48	.140
	Gender	-3.91	-0.27	1.22	-3.22	.002
	FinTime	-0.01	-0.38	0.00	-3.85	< .001

Note. Demographic predictors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), and FinTime to finish time.

Figure 2.2

Scatter plots illustrating the relationships between demographic factors and realistic prediction bias and absolute accuracy.



Note. Panel A shows the relationship between finish time and bias. Panel B shows the relationship between training volume and bias. Panel C illustrates the relationship between gender and bias relative to finish time. Panel D shows the relationship between training volume and absolute accuracy.

Absolute accuracy. Results from correlational analyses and the multiple regression models for realistic prediction absolute accuracy can be seen in Tables 2.6 and 2.8. Neither regression model was significant in predicting absolute accuracy regardless of whether finish time was included as a factor, $R^2 = .053$, $F_{(5, 182)} = 2.02$, $p = .078$, or not, $R^2 = .055$, $F_{(6, 181)} = 1.76$, $p = .109$. Training volume was a significant predictor of absolute accuracy in correlational analyses and the regression model that did not include finish time, with participants who ran more kilometres per week being more precise than participants who ran fewer kilometres (Fig. 2.2D). However, training volume's contribution was non-significant in the model that included finish time. This is likely because training volume and finish time share variance, and finish time accounted for some of this variance in the regression model. Overall, the only factor with the capacity to predict realistic prediction accuracy was training volume, as the rest of the factors failed to reach significance in either model.

Table 2.8

Multiple regression coefficients for demographic predictors on the outcome variable realistic prediction absolute accuracy.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No finish time	Intercept	9.61	—	1.55	6.21	< .001
	km	-0.08	-0.20	0.03	-2.55	.011
	Months	0.00	-0.04	0.00	-0.46	.644
	Club	0.59	0.04	1.06	0.56	.576
	Age	-0.05	-0.10	0.04	-1.32	.188
	Gender	0.53	0.05	0.81	0.65	.514
Finish time included	Intercept	6.93	—	4.09	1.70	.092
	km	-0.07	-0.17	0.03	-1.94	.054
	Months	0.00	-0.02	0.00	-0.28	.776
	Club	0.64	0.05	1.06	0.61	.545
	Age	-0.06	-0.12	0.04	-1.45	.149
	Gender	0.86	0.08	0.94	0.92	.359
	FinTime	0.00	0.07	0.00	0.71	.479

Note. Demographic predictors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), and FinTime to finish time.

2.2.3.4.2 Goal Predictions

Bias. Results from correlational analyses and the multiple regression models for goal prediction bias can be seen in Tables 2.6 and 2.9. The model that did not include finish time was not significant in predicting goal prediction bias, $R^2 = .025$, $F_{(5, 186)} = 0.95$, $p = .452$. In contrast, the model that included finish time had a significant contribution to goal prediction bias, $R^2 = .102$, $F_{(6, 185)} = 3.52$, $p = .003$. Runners who were faster to finish the race were more likely to be underconfident, whilst slower runners were more likely to be overconfident (Fig. 2.3A). After accounting for finish time, older runners were significantly more underconfident or less overconfident than younger runners (Fig. 2.3B). Similarly, female runners tended to exhibit higher underconfidence in their goal predictions than male runners after including finish time in the regression model. As illustrated in Figure 2.3C, this difference was likely driven by higher female relative to male underconfidence for faster and middle finish times, whereas participants of either gender exhibited more similar bias for slower finish times. Training volume, months of running experience, and club membership were not significant predictors of goal prediction bias in either model.

Table 2.9

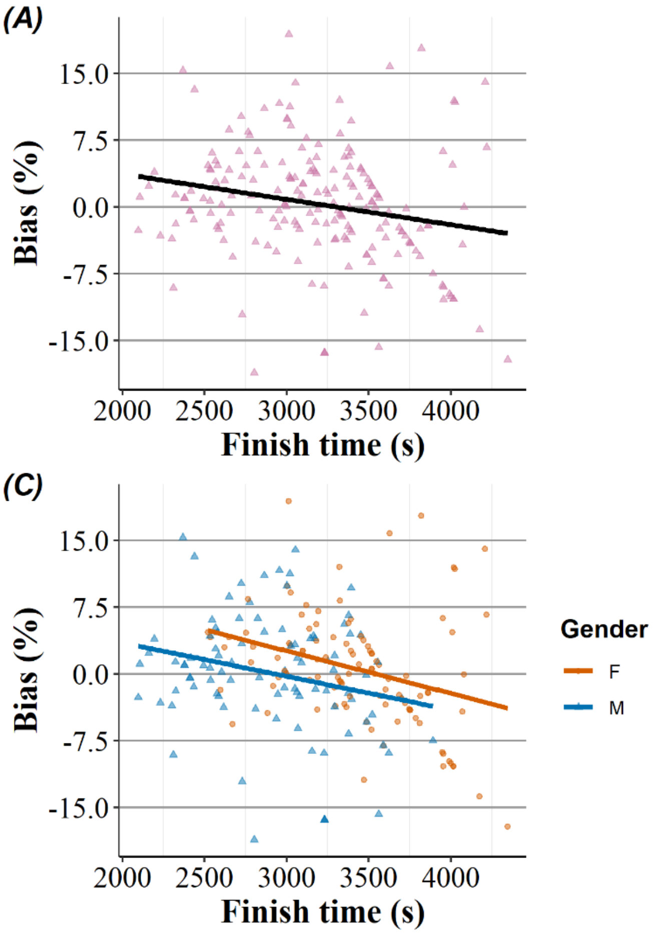
Multiple regression coefficients for demographic predictors on the outcome variable goal prediction bias.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No finish time	Intercept	-2.47	—	1.85	-1.33	.185
	km	0.03	0.06	0.04	0.72	.472
	Months	0.00	0.01	0.00	0.16	.876
	Club	1.15	0.07	1.27	0.91	.366
	Age	0.06	0.10	0.05	1.28	.204
	Gender	-0.51	-0.04	0.98	-0.52	.603
Finish time included	Intercept	15.10	—	4.74	3.19	< .001
	km	-0.05	-0.10	0.04	-1.17	.242
	Months	0.00	-0.06	0.00	-0.77	.440
	Club	0.78	0.05	1.22	0.64	.526
	Age	0.11	0.18	0.05	2.26	.025
	Gender	-2.72	-0.21	1.09	-2.49	.014
	FinTime	-0.01	-0.39	0.00	-4.00	< .001

Note. Demographic predictors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), and FinTime to finish time.

Figure 2.3

Scatter plots illustrating the relationships between demographic factors and goal prediction bias.



Note. Panel A shows the relationship between finish time and bias. Panel B shows the relationship between age and bias. Panel C illustrates the relationship between gender and bias relative to finish time.

Absolute accuracy. Results from correlational analyses and the multiple regression models for goal prediction absolute accuracy can be seen in Tables 2.6 and 2.10. Both models were significant in predicting goal prediction absolute accuracy, regardless of whether finish time was included, $R^2 = .115$, $F_{(6,\,185)} = 4.01$, $p = .001$, or not, $R^2 = .065$, $F_{(5,\,186)} = 2.59$, $p = .027$. However, the R^2 values suggest that the model that included finish time had higher capacity to explain absolute accuracy variance than the model that did not include finish time. Though higher training volume was predictive of higher precision in the model without performance (Fig. 2.4A), and exhibited a negative correlation with absolute accuracy when examined individually, it was not a significant predictor of precision in the model that included finish time. This suggests that finish time accounted for training volume's capacity to explain absolute accuracy variance. Finish time in itself was a significant predictor of absolute accuracy, with slower runners being less precise in their goal estimates than faster runners (Fig. 2.4B). Age was a significant factor in both models, as older runners were more precise than younger runners (Fig. 2.4C). In the model accounting for finish time, gender was a significant predictor of precision, with male runners being less precise than female runners. Figure 2.4D shows that this difference was not present for slower finish times, so it was likely driven by faster finish times. Months of running experience and club membership did not contribute to absolute accuracy in either model.

Table 2.10

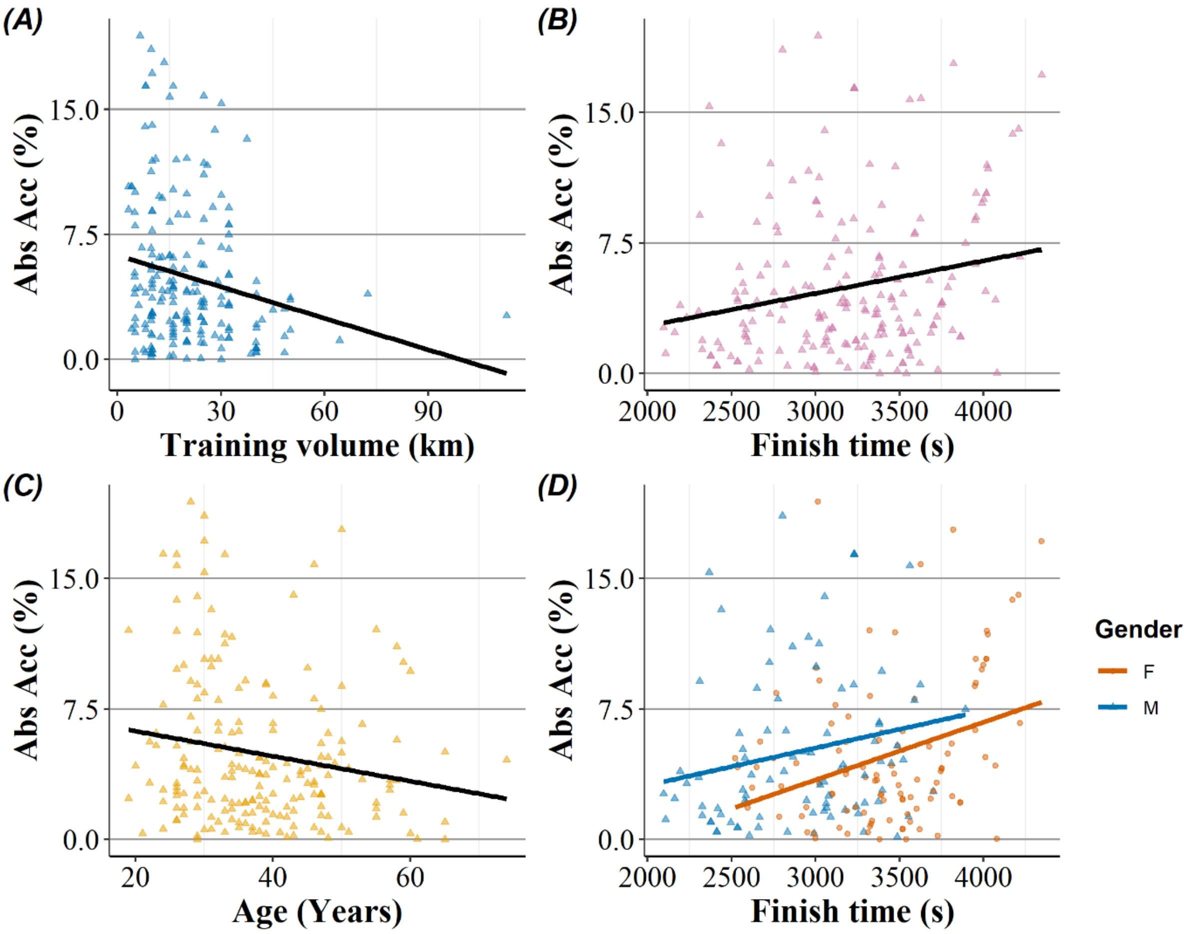
Multiple regression coefficients for demographic predictors on the outcome variable goal prediction absolute accuracy.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No finish time	Intercept	8.28	—	1.20	6.92	< .001
	km	-0.06	-0.21	0.02	-2.69	.008
	Months	0.00	0.02	0.00	0.26	.796
	Club	0.34	0.03	0.82	0.42	.677
	Age	-0.07	-0.16	0.03	-2.11	.036
	Gender	0.56	0.07	0.63	0.88	.380
Finish time included	Intercept	-1.02	—	3.11	-0.33	.744
	km	-0.03	-0.08	0.03	-0.97	.335
	Months	0.00	0.08	0.00	1.01	.314
	Club	0.54	0.05	0.80	0.67	.503
	Age	-0.09	-0.22	0.03	-2.88	.004
	Gender	1.73	0.20	0.72	2.41	.017
	FinTime	0.00	0.32	0.00	3.23	.001

Note. Demographic predictors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), and FinTime refers to finish time.

Figure 2.4

Scatter plots illustrating the relationships between demographic factors and goal prediction absolute accuracy.



Note. Panel A shows the relationship between training volume and absolute accuracy. Panel B shows the relationship between finish time and absolute accuracy. Panel C illustrates the relationship between age and absolute accuracy. Panel D shows the relationship between gender and absolute accuracy relative to finish time.

2.2.4 Discussion of Study 1

In Study 1, I examined the relationships of expertise, experience, age, and gender with calibration using realistic and goal performance predictions made within 24 hours before the Edinburgh Christmas 10k Run.

Prediction type results were surprising as, contrary to my expectations of goal predictions being overconfident and realistic predictions being less overconfident, goal predictions were unbiased, whilst realistic predictions were underconfident (~3.9%) and less precise than goal predictions. Given that, to my knowledge, this was the first study to explore the effects of prediction type on calibration, it is important to implement the present design in other running competitions to test the replicability of these findings.

I assessed expertise in terms of finish time, and expected faster runners to be underconfident and precise, and slower runners to be overconfident and less precise than faster runners. In line with these predictions, slower finish times were associated with decreases in underconfidence in realistic predictions, though performance did not predict realistic prediction absolute accuracy. In goal predictions, faster finish times predicted underconfidence, whilst slower finish times predicted overconfidence or lower underconfidence. Slower finish times were also predictive of lower precision than faster finish times. Therefore, my results were largely in line with my predictions for the relationship between expertise and calibration.

I predicted that participants with higher experience would be better calibrated than participants with lower experience. Results on the relationship between experience markers and calibration provided mixed support for these predictions. Months of running experience and club membership were not associated with calibration for either prediction type. In fact, club membership did not even predict performance estimates and finish time after accounting for other experience factors—either because it is a secondary experience factor or because the sample size of club members was too small (n = 38) to detect the factor's influence. Conversely, higher training volume was predictive of lower absolute accuracy, and thus higher precision, in both realistic and goal predictions. However, its contribution was limited or non-existent after accounting for finish time. This is not surprising, as training volume and performance are closely connected (see Section 2.2.3.2). Therefore, though training volume contributes to prediction precision, finish time accounts for this contribution. It is thus likely that finish time is a better predictor of calibration than other experience factors. Nonetheless, further research from races of different lengths is required to examine whether finish time accounts for the contribution of other experience factors consistently.

I anticipated that older age would predict better calibration than younger age, and was interested in the extent to which this relationship would be present after accounting for other experience factors. Age showed a dissociation in its relationship with calibration between realistic and goal predictions. Though not associated with either bias or absolute accuracy in realistic predictions, it supported my prediction and had a significant contribution to both in goal predictions. Older runners were more underconfident but also more precise in their goal predictions than younger runners. Interestingly, these relationships remained present after accounting for finish time and other experience factors, suggesting that age had an independent influence on goal prediction calibration.

I was also interested in the extent to which male runners would exhibit higher overconfidence than female runners in line with research in longer races (e.g. marathons; see Section 1.3.1.4). Supporting long distance studies, gender was associated with calibration, but only after accounting for finish time variance. In realistic predictions, the fastest male and female runners appeared to be similarly underconfident, but male runners were less underconfident for slower finish times than female runners, who were consistently underconfident. There were no gender differences in precision for realistic predictions. In goal predictions, slower runners of either gender appeared to be similarly overconfident, but faster female runners were more underconfident than male runners. Surprisingly, faster male runners showed lower precision in their goal predictions than faster female runners, though this difference was not present for slower runners. These results support previous findings of relative female underconfidence (Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018), though gender's influence on calibration appears to depend on finish time. Interestingly, I observed patterns of lower male underconfidence, rather than higher male overconfidence, but this can be attributed to the general lack of overconfidence in the sample.

Overall, Study 1 produced important results on the relationships between demographic factors and calibration. Additionally, prediction type had a significant effect on calibration, indicating the need to take it in consideration when examining prediction accuracy in athletic events. Nonetheless, some findings were not in the anticipated direction (e.g. experience), and data from only one race are not sufficient to fully understand the nature of the relationships examined. For these reasons, it was important to further investigate factors associated with running calibration by examining other races and by using different data collection methods.

2.3 STUDY 2 – ALLOA HALF MARATHON

2.3.1 Study specifics

In Study 2, I collected demographic and calibration data from the 2019 Alloa Half Marathon, which took place on Sunday, 31st of March in the town of Alloa in Clackmannanshire, Scotland. It was organised by the Alloa Round Table organisation and covered a distance of 13.1 miles, and had 1816 finishers. It is a relatively flat route (~100m elevation gain) and can be influenced by weather conditions, though the weather on the day of the race was good, by Scottish standard, for running—approximately 8°C, 7km/h winds (gusting to 8km/h), no rain. It is one of the first road half marathon in the annual road-running season, and attracts recreational runners of a wide range of abilities.

The data collected were similar data to Study 1, as participants were asked to provide their demographic information (i.e. training volume, months of running experience, club membership, age, and gender), and realistic and goal predictions. My predictions for Study 2 were consistent with my predictions for Study 1 and the overall predictions I presented at the end of Section 2.1. What was different from Study 1, where I collected data in person and within 24 hours before the race, was that runners in Study 2 completed online questionnaires at any point during the 150 days leading to the race. This allowed me to also assess the relationship between prediction lead (i.e. number of days before the race when predictions were made) and calibration. I expected runners to make more accurate performance estimates closer to the time of the race, when they should be better aware of their athletic capacity to perform. Furthermore, it was important to examine whether results from Study 1 would be replicated using data collected online, as the physical presence of researchers in Study 1 could have influenced how participants filled in information and made predictions. Finally, a key difference between the Study 2 and Study 1 is the distance examined. A half marathon is a more challenging distance to compete in than 10km, so the Alloa Half Marathon may draw on a different type of runner than the Edinburgh Christmas 10k Run (i.e. more competitive runners that do not participate just "for fun").

2.3.2 Methods

2.3.2.1 Participants

I collected finish time predictions online from 402 runners who participated in the 2019 Alloa Half Marathon. There were 25 duplicate sets where participants had entered their predictions twice. From these sets, I only analysed the first predictions input (though data from one runner were removed completely, as the two duplicate predictions were different by over an hour), because I wanted to have a higher number of predictions that were made further in advance before the race to ensure higher power for prediction lead analyses. I then matched the remaining 376 predictions to the corresponding finish times. Sixty-three runners' predictions could not be matched (they either did not participate or finish the race). This left 313 data-points considered in the next stage of the analysis.

I then removed outliers based on absolute accuracy percentages (relative to finish time) to limit the influence of extreme predictions in the analyses (z-scores with absolute values greater than three). I removed seven data-points (five female and two male runners; seven unaffiliated runners) from realistic prediction analyses, leading to 306 matched data-points (137 women and 169 men; 97 club members and 209 unaffiliated runners; Age: $M = 42.4$ years, $SD = 10.0$ years). Similarly, I removed seven data-points (four female and three male runners; one club and six unaffiliated runners) from goal analyses, leading to 306 data-points (138 women and 168 men; 96 club members and 210 unaffiliated runners; Age: $M = 42.3$ years, $SD = 10.0$ years). For prediction type analyses, I removed ten outliers (six female and four male runners; one club and nine unaffiliated runners) from both realistic and goal predictions, producing 303 data-points (136 women and 167 men; 96 club members and 207 unaffiliated runners; Age: $M = 42.4$ years, $SD = 10.0$ years). Overall, after removing four data-points (three female and one male runners; four unaffiliated runners) that were outliers in both prediction types, my total sample for finish time analyses consisted of 309 data-points (139 women and 170 men; 97 club members and 212 unaffiliated runners; Age: $M = 42.3$ years, $SD = 10.0$ years).

The present study received ethical approval from the University of St Andrews School of Psychology & Neuroscience Ethics Committee (Ethics approval code: PS13876; see Appendix 8.1.2). Participants had the option to enter their email address in a draw for a prize of 3 x £20 and 5 x £10 Amazon vouchers. The draw took place following the race.

2.3.2.2 Materials

To participate in the Alloa Half Marathon, athletes registered online. I set up an online questionnaire (see Appendix 8.3), linked to from the official race registration portal and online running fora, using which runners could input their demographic information (training volume, years/months of running experience, and age). The website also requested that runners indicated their goal (i.e. "The finish time I hope to achieve (my goal time) is:") and realistic (i.e. "The finish time I think is most likely for me to achieve is:") predictions. The questionnaire asked participants to read instructions for both prediction types prior to completing this section. Upon completion, the date when the participants filled in the questionnaire was recorded to produce prediction lead data.

2.3.2.3 Design

The design and analyses I implemented in Study 2 were similar to those of Study 1. The aim of the study was thus to examine the influence of expertise (i.e. finish time), experience (i.e. training volume, months of running experience, and club membership), age, gender, prediction type (goal vs realistic predictions), and prediction lead (i.e. number of days before the race when the predictions were made) on bias and absolute accuracy. Akin to study 1, I asked participants to read instructions for both prediction types prior to completion to control for the order in which the instructions were delivered. Additionally, I calculated bias and absolute accuracy percentages relative to finish time to account for the effects of performance value variation on calibration.

2.3.2.4 Procedure

Runners who registered for the 2019 Alloa Half Marathon were given the option of participating in the study by following a link to the study website. There, they gave informed consent and provided the data required, as outlined above. I collected data until the 31st of March, when the Alloa Half Marathon took place. Following the race, I matched the questionnaire data with published finish times, and participant gender and club membership.

2.3.3 Results

2.3.3.1 Data checks

Consistent with the data checks carried out in Study 1 (see Section 2.2.3.1), I compared finish times from the 2016, 2017, 2018, and 2019 Alloa Half Marathons to examine whether performance in the 2019 race investigated here was similar to performance in the same course in different years. I then compared finish times between runners who participated in the study and runners who did not, to test whether participants had a different profile to non-participants. Finally, I examined prediction type frequencies to test whether participants were more likely to make faster goal or realistic predictions.

Differences in finish time between the 2016, 2017, 2018, and 2019 Alloa Half Marathons. The mean finish time of the 2019 race examined in the present study was 115 minutes (*SD* = 21 min). The mean finish time of the 2018 race was 120 minutes (*SD* = 22 min), of the 2017 race was 115 minutes (*SD* = 21 min), and of the 2016 race was 114 minutes (*SD* = 21 min). A one-way between subjects ANOVA showed a significant difference between the finish times of the four races, $F_{(3, 8434)} = 34.91$, $p < .001$, $\eta p^2 = .012$. Pairwise comparisons using a Sidak correction showed that the 2018 race was significantly slower than all three other races (all *p*s < .001). In contrast, the average finish time of the 2019 race examined was similar to the 2017 and 2016 races (all *p*s > .10), suggesting it was consistent with general performance patterns in the Alloa Half Marathon.

Differences in finish time between sample and non-participants. The mean finish time of the 1501 runners who finished the 10km race but did not participate in the study was 105 minutes (*SD* = 21 min). The average finish time of the 309 runners comprising the present sample (~17% of the field of 1816 runners) was 105 minutes (*SD* = 22 min), indicating that runners in the sample performed similarly to runners who did not participate in the study $t_{(426.6)} = 0.11$, $p = .913$, $d = 0.01$.

Prediction type frequencies. Of the 303 participants who provided valid realistic and goal predictions, only seven participants (~2%) made realistic predictions that were faster than their goal predictions; forty-four (~15%) made the same realistic and goal predictions; and two-hundred and fifty-two (~83%) made goal predictions that were faster than their realistic predictions. A chi-squared test indicated that there was a significant difference in the number of participants who made goal predictions that were faster, equal, or slower compared to realistic predictions, $\chi^2_{(2)} = 345.41$, $p < .001$. The majority of runners made goal predictions that were faster than their realistic predictions, whilst only a small minority made goal predictions that matched or were slower than their realistic predictions.

2.3.3.2 Performance & predictions

I examined the capacity of experience, age, gender, and prediction lead to predict performance, realistic predictions, and goal predictions using the same methodology as Study 1, which I described in Section 2.2.3.2.

Table 2.11

Descriptive statistics for Club membership.

	Club		Unaffiliated	
Outcome Variable	***Mean***	***SD***	***Mean***	***SD***
Finish Time	111 min	23 min	116 min	21 min
Realistic Prediction	112 min	23 min	118 min	21 min
Goal Prediction	107 min	20 min	113 min	19 min
Realistic Bias	0.31%	6.33%	1.80%	6.30%
Realistic Absolute Accuracy	4.94%	3.93%	4.79%	4.45%
Goal Bias	-3.36%	5.72%	-2.08%	6.10%
Goal Absolute Accuracy	4.94%	4.41%	4.76%	4.34%

Note. The table provides information on the means and standard deviations (SD) of club members and unaffiliated runners for performance, predictions, and calibration.

Table 2.12

Descriptive statistics for Gender.

	Male		Female	
Outcome Variable	***Mean***	***SD***	***Mean***	***SD***
Finish Time	105 min	19 min	126 min	20 min
Realistic Prediction	107 min	18 min	127 min	20 min
Goal Prediction	103 min	17 min	121 min	18 min
Realistic Bias	1.36%	6.20%	1.29%	6.52%
Realistic Absolute Accuracy	4.59%	4.37%	5.15%	4.18%
Goal Bias	-1.93%	5.98%	-3.16%	5.98%
Goal Absolute Accuracy	4.57%	4.30%	5.12%	4.41%

Note. The table provides information on the means and standard deviations (SD) of club members and unaffiliated runners for performance, predictions, and calibration.

Finish time. Results from correlational analyses and the multiple regression model for performance can be seen in Tables 2.13 and 2.14. The regression model was significant in predicting finish time, $R^2 = .419$, $F_{(6, 301)} = 36.22$, $p < .001$. All predictors other than club membership and prediction lead were significant contributors to finish time. Runners who engaged in higher training volume and had been running for longer were faster to finish the race than runners who engaged in lower training volume and had been running for a shorter period. Club members only showed a non-significant tendency to complete the race faster than unaffiliated runners (Table 2.11), whilst older and female runners were significantly slower to finish the race than younger and male runners respectively (Table 2.12). Prediction lead did not exhibit a significant relationship with finish time. Regression results were similar with the individual correlational associations between demographic factors and finish time.

Table 2.13

Correlation coefficients for the associations of demographic factors and prediction lead with the outcome variables finish time, realistic predictions, and goal predictions.

	Performance		**Realistic**		**Goal**	
Factor	***r***	***p***	***r***	***p***	***r***	***p***
km	-0.47	< .001	-0.53	< .001	-0.50	< .001
Months	-0.21	< .001	-0.26	< .001	-0.25	< .001
Club	-0.10	.091	-0.13	.019	-0.14	.016
Age	0.11	.048	0.12	.045	0.10	.084
Gender	-0.47	< .001	-0.48	< .001	-0.47	< .001
PredLead	0.07	.231	0.04	.511	0.05	.343

Note. *r* represents the correlation coefficient of each factor with the outcome variables. *p* represents the *p*-value associated with corresponding predictor and outcome variable. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), and PredLead to prediction lead.

Table 2.14

Multiple regression coefficients for demographic and prediction lead predictors on the outcome variable finish time.

Coefficient	*B*	*Beta*	Std. Error	*t*	*p*
Intercept	120.03	—	4.55	26.38	< .001
km	-0.35	-0.34	0.05	-6.99	< .001
Months	-0.03	-0.18	0.01	-3.70	< .001
Club	-3.31	-0.07	2.16	-1.53	.127
Age	0.51	0.24	0.10	4.94	< .001
Gender	-17.93	-0.41	2.02	-8.90	< .001
PredLead	0.00	0.01	0.02	0.19	.851

Note. Demographic predictors were entered at the same time in the multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), and PredLead to prediction lead.

Predictions. Results from correlational analyses and the multiple regression models for realistic and goal predictions can be seen in Tables 2.13 and 2.15. The variance explained by each regression model was significant: realistic predictions—R^2 = .494, $F_{(6, 298)} = 48.49$, $p <$ .001; goal predictions—R^2 = .458, $F_{(6, 298)} = 41.89$, $p < .001$. Regardless of prediction type, runners with more experience (i.e. higher training volume, more months of running experience, and club members—Table 2.11) made significantly faster predictions than runners with less experience (i.e. lower training volume, fewer months of running experience, and unaffiliated members). Older runners were significantly more likely to make slower realistic and goal realistic predictions than younger runners. This tendency was non-significant for goal predictions in correlational analyses, but became significant in the multiple regression model after accounting for shared variance. Male runners were more likely to make faster predictions than female runners for both prediction types (Table 2.12). Prediction lead was not a significant contributor for either prediction type. Multiple regression results for finish time predictions were largely in line with finish time results, though club membership exhibited only a non-significant tendency to predict performance.

Table 2.15

Multiple regression coefficients for demographic and prediction lead predictors on the outcome variables realistic and goal predictions.

Prediction type	**Coefficient**	***B***	***Beta***	***Std. Error***	***t***	***p***
Realistic	Intercept	122.90	—	4.20	29.29	< .001
	km	-0.38	-0.37	0.05	-8.21	< .001
	Months	-0.04	-0.23	0.01	-5.02	< .001
	Club	-4.83	-0.10	1.99	-2.43	.016
	Age	0.55	0.26	0.10	5.75	< .001
	Gender	-17.91	-0.41	1.86	-9.65	< .001
	PredLead	-0.01	-0.02	0.02	-0.52	.603
Goal	Intercept	117.58	—	3.99	29.49	< .001
	km	-0.32	-0.35	0.04	-7.39	< .001
	Months	-0.04	-0.22	0.01	-4.66	< .001
	Club	-4.77	-0.11	1.90	-2.52	.012
	Age	0.47	0.24	0.09	5.18	< .001
	Gender	-16.19	-0.41	1.77	-9.17	< .001
	PredLead	0.00	-0.01	0.02	-0.11	.915

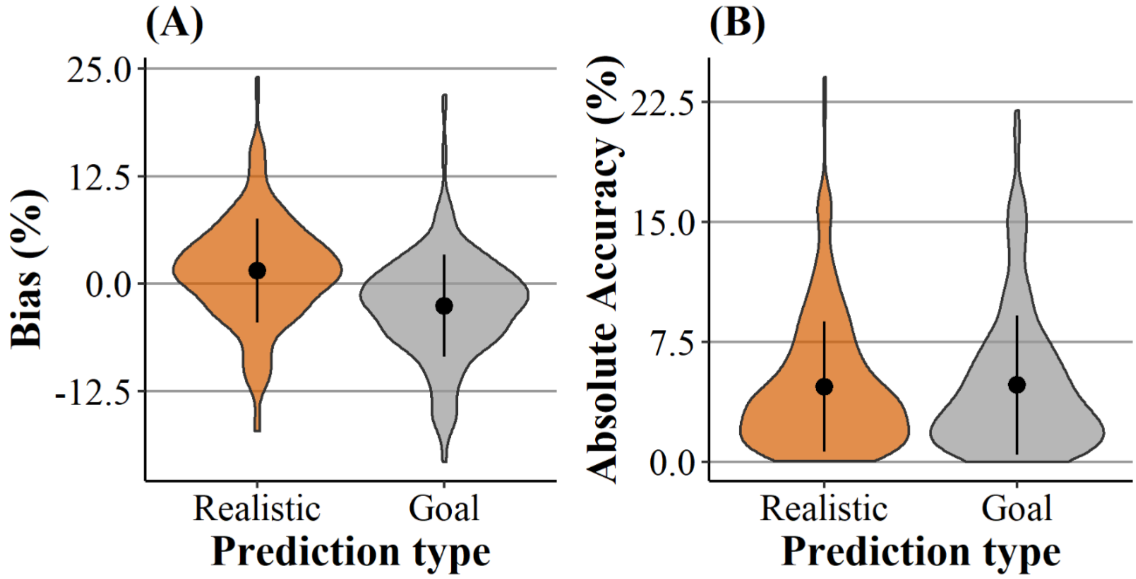
Note. Demographic predictors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), and PredLead to prediction lead.

2.3.3.3 Calibration – Prediction type

Bias. To test differences in bias between goal and realistic predictions, I conducted a paired samples t-test, which indicated a significant difference between them, $t_{(302)} = 18.93$, $p < .001$, $d = 1.09$. Two one-sample t-tests comparing realistic and goal prediction bias to zero, i.e. no bias, showed that realistic predictions were significantly underconfident ($M = 1.53\%$, $SD = 6.03\%$), $t_{(302)} = 4.41$, $p < .001$, $d = 0.25$, whilst goal predictions were significantly overconfident ($M = -2.56\%$, $SD = 5.95\%$), $t_{(302)} = -7.50$, $p < .001$, $d = -0.43$ (Fig. 2.5A).

Figure 2.5

Violin plots illustrating the effects of prediction type on bias and absolute accuracy.



Note. Panel A illustrates prediction type effects on bias. Panel B shows prediction type effects on absolute accuracy. The perimeter of each violin plot illustrates density, the central point represents the mean, and the vertical line represents +/- one standard deviation.

Absolute accuracy. To test differences in absolute accuracy between goal and realistic predictions, I conducted a paired samples t-test, which did not show a significant difference between them, $t_{(302)} = 0.38$, $p = .702$, $d = 0.02$ (Fig. 2.5B). Goal absolute accuracy (M = 4.79%, SD = 4.35%) was similar to realistic absolute accuracy (M = 4.70%, SD = 4.06%). Overall, participants were underconfident and overconfident in their realistic and goal predictions respectively, but similarly precise.

2.3.3.4 Calibration – Demographic factors

2.3.3.4.1 Realistic Predictions

Bias. I examined the capacity of expertise, experience, age, gender, and prediction lead to predict calibration using the same methodology as in Study 1, which I described in Section 2.2.3.4.

Table 2.16

Correlation coefficients for the associations of demographic factors and prediction lead with the outcome variables bias and absolute accuracy for realistic and goal predictions.

	Real Bias		**Real Abs Acc**		**Goal Bias**		**Goal Abs Acc**	
Factor	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***
km	-0.16	.005	-0.12	.030	-0.03	.624	-0.08	.163
Months	-0.15	.011	-0.14	.013	-0.09	.100	-0.03	.628
Club	-0.11	.056	0.02	.778	-0.10	.085	0.02	.740
Age	0.00	1.000	0.01	.884	-0.05	.412	0.00	.941
Gender	0.01	.925	-0.07	.254	0.10	.073	-0.06	.271
PredLead	-0.09	.123	0.13	.029	-0.04	.530	0.14	.018
FinTime	-0.21	$< .001$	0.20	$< .001$	-0.33	$< .001$	0.34	$< .001$

Note. *r* represents the correlation coefficient of each factor with the outcome variables. *p* represents the *p*-value associated with corresponding predictor and outcome variable. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), PredLead to prediction lead, and FinTime to finish time.

Results from correlational analyses and the multiple regression models for realistic prediction bias can be seen in Tables 2.16 and 2.17. Both regression models were significant in predicting bias regardless of whether they included finish time, $R^2 = .206$, $F_{(7, 297)} = 11.02$, $p < .001$, or not, $R^2 = .058$, $F_{(6, 298)} = 3.07$, $p = .006$. Nonetheless, the model that included finish explained more bias variance than the model that did not include finish time. In the latter model, only training volume and months of running experience were significant predictors of bias. In the model that included finish time, all factors, other than prediction lead, were significant predictors. Faster finish times were associated with underconfidence, whereas slower finish times were associated with overconfidence (Fig. 2.6F). Higher training volume and more months of running experience were linked with a decrease in underconfidence and an increase in overconfidence across models (Fig. 2.6A & 2.6B). Club members were less underconfident than unaffiliated runners for faster, but not slower, finish times, as indicated by Figure 2.6C. After accounting for variance explained by finish time, age became a significant contributor of bias, with older runners being more likely to be underconfident or less overconfident than younger runners (Fig. 2.6D). Significant gender differences in bias also arose after controlling for finish time, with male runners showing a tendency to be less underconfident than female runners for faster, but not slower, finish times, as illustrated by Figure 2.6E. Overall, accounting for finish time variance in the second model allowed me to observe relationships between demographic factors and realistic prediction bias that were not otherwise visible in the model that did not include finish time and in correlational analyses.

Table 2.17

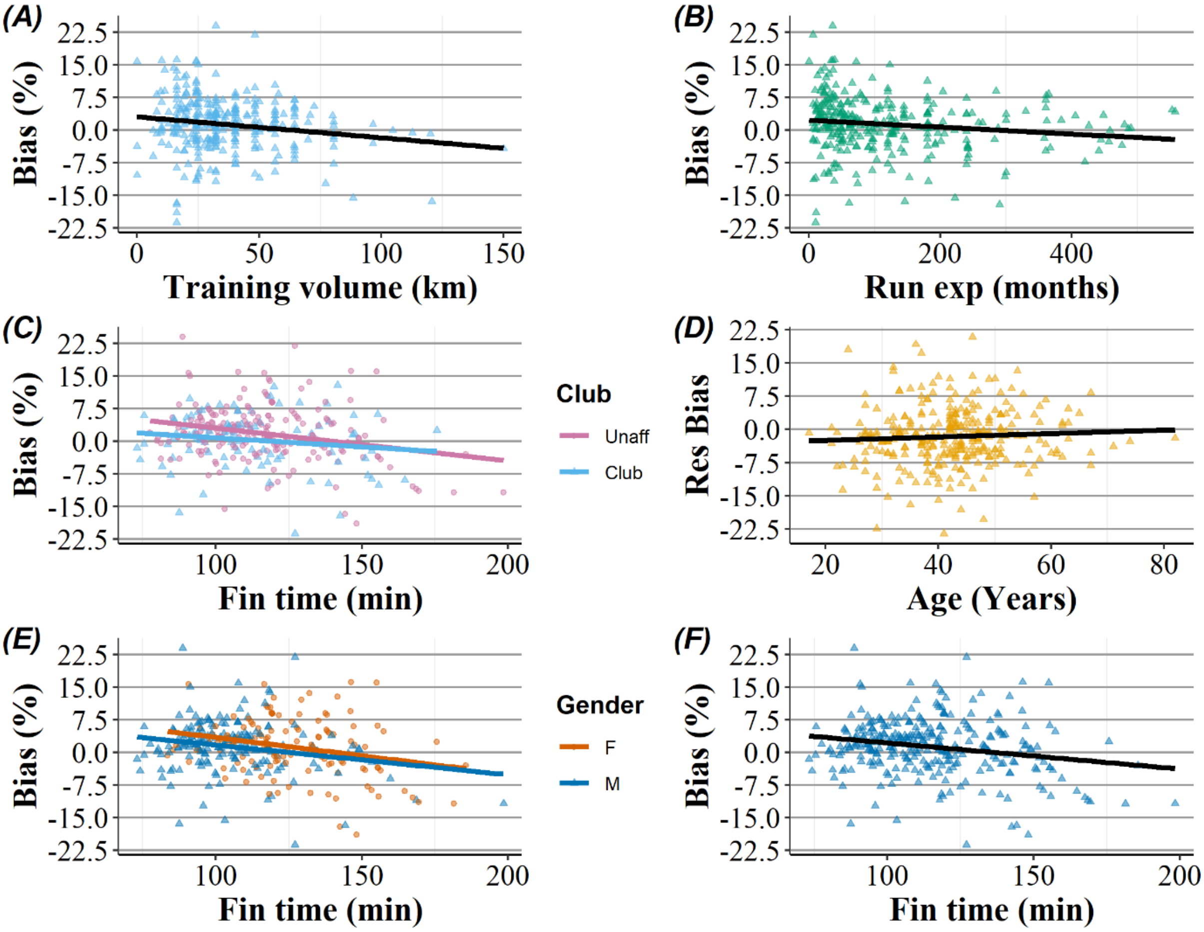
Multiple regression coefficients for demographic and prediction lead predictors on the outcome variable realistic prediction bias.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No finish time	Intercept	2.63	—	1.69	1.56	.120
	km	-0.04	-0.13	0.02	-2.03	.043
	Months	-0.01	-0.14	0.00	-2.20	.029
	Club	-1.18	-0.09	0.80	-1.48	.141
	Age	0.04	0.07	0.04	1.07	.286
	Gender	0.34	0.03	0.75	0.45	.654
	PredLead	-0.01	-0.10	0.01	-1.67	.097
Finish time included	Intercept	20.15	—	2.82	7.15	< .001
	km	-0.09	-0.29	0.02	-4.81	< .001
	Months	-0.01	-0.23	0.00	-3.89	< .001
	Club	-1.68	-0.12	0.74	-2.82	.023
	Age	0.12	0.18	0.04	3.16	.002
	Gender	-2.30	-0.18	0.77	-2.98	.003
	PredLead	-0.01	-0.09	0.01	-1.69	.093
	FinTime	-0.15	-0.51	0.02	-7.45	< .001

Note. Demographic predictors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), PredLead to prediction lead, and FinTime to finish time.

Figure 2.6

Scatter plots illustrating the relationships between demographic factors and realistic prediction bias.



Note. Panel A shows the relationship between training volume and bias; panel B between months of running experience and bias; panel C between club membership and bias relative to finish time; panel D between age and residual bias (i.e. bias minus the regression intercept and the finish time variance related to bias); panel E between gender and bias relative to finish time; and panel F between finish time and bias.

Absolute accuracy. Results from correlational analyses and the multiple regression models on realistic prediction absolute accuracy can be seen in Tables 2.16 and 2.18. Both models were significant in predicting absolute accuracy regardless of whether they included finish time, $R^2 = .068$, $F_{(7, 297)} = 3.10$, $p = .004$, or not, $R^2 = .049$, $F_{(6, 298)} = 2.56$, $p = .020$. Nonetheless, the model that included finish time explained more absolute accuracy variance than the model that did not. Contrary to correlational analyses where training volume exhibited a negative correlation with absolute accuracy, training volume was not a significant predictor of precision in either regression model, as it contributed little unique variance beyond the shared variance accounted for by months of running experience and performance (Fig. 2.7A). Months of running experience exhibited a significant negative correlation with absolute accuracy and was a significant predictor in the model that did not include finish time, with more experienced runners being more precise in their predictions than less experienced runners (Fig. 2.7B). However, this association was non-significant in the model that included finish time. Finish time in itself was a significant predictor of absolute accuracy, with faster runners showing higher precision than slower runners (Fig. 2.7D). Finally, prediction lead was a significant contributor to absolute accuracy in both models, with runners who made their predictions closer to the time of the race being more precise than runners making their predictions earlier (Fig. 2.7C). Club membership, age, and gender did not contribute to realistic prediction absolute accuracy in either model—nor did they correlate with it when examined individually.

Table 2.18

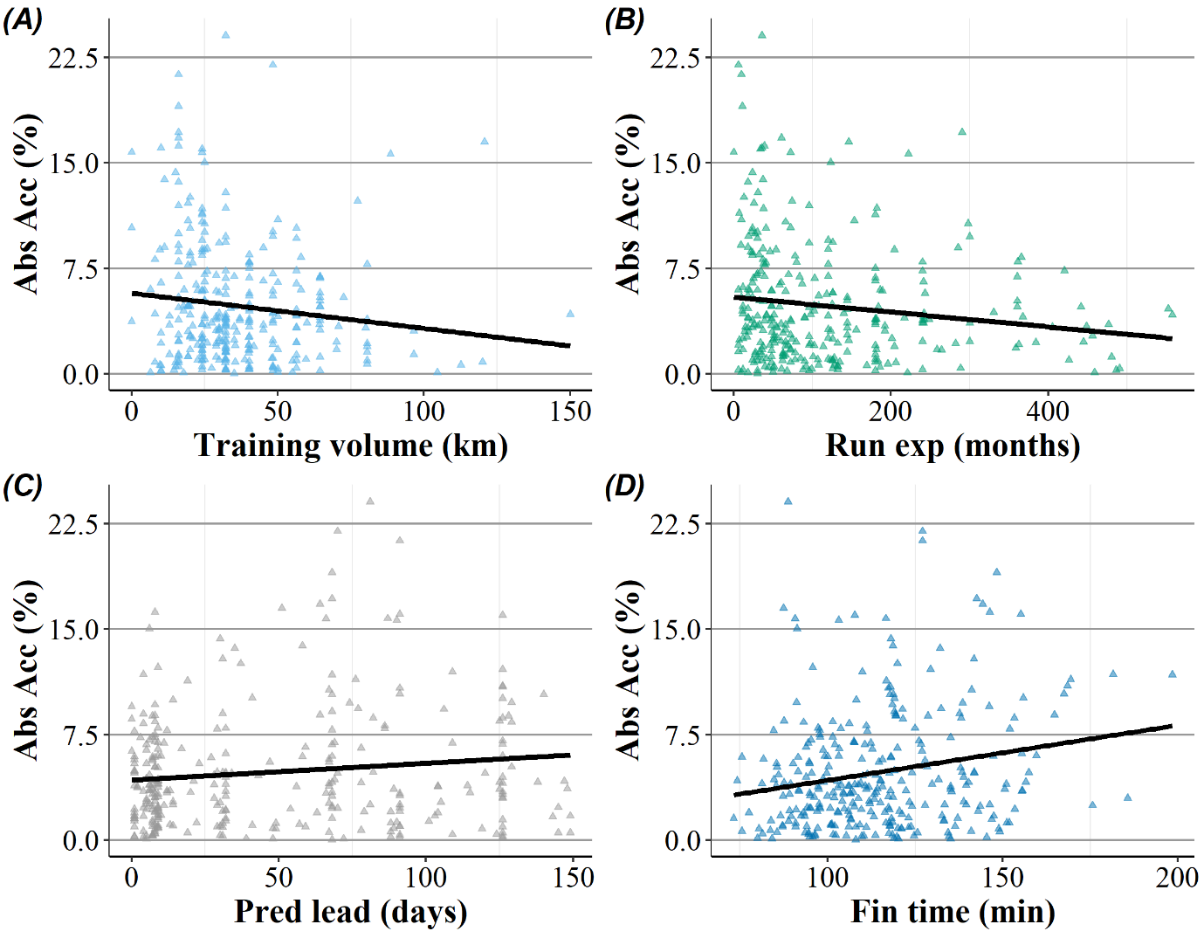
Multiple regression coefficients for demographic and prediction lead predictors on the outcome variable realistic prediction absolute accuracy.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No finish time	Intercept	4.77	—	1.15	4.16	< .001
	km	-0.02	-0.08	0.01	-1.36	.175
	Months	-0.01	-0.14	0.00	-2.29	.022
	Club	0.28	0.03	0.54	0.51	.611
	Age	0.02	0.05	0.03	0.77	.440
	Gender	-0.31	-0.04	0.51	-0.61	.544
	PredLead	0.01	0.12	0.01	2.10	.037
Finish time included	Intercept	0.52	—	2.07	0.25	.802
	km	-0.01	-0.02	0.01	-0.36	.717
	Months	0.00	-0.11	0.00	-1.75	.081
	Club	0.40	0.04	0.54	0.74	.461
	Age	0.00	0.01	0.03	0.07	.941
	Gender	0.33	0.04	0.57	0.59	.558
	PredLead	0.01	0.12	0.01	2.07	.039
	FinTime	0.04	0.18	0.01	2.46	.014

Note. Demographic predictors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), PredLead to prediction lead, and FinTime to finish time.

Figure 2.7

Scatter plots illustrating the relationships between demographic factors and realistic prediction absolute accuracy.



Note. Panel A shows the relationship between training volume and absolute accuracy; panel B between months of running experience and absolute accuracy; panel C between prediction lead and absolute accuracy; and panel D between finish time and absolute accuracy.

2.3.3.4.2 Goal Predictions

Bias. Results from correlational analyses and the multiple regression models for goal prediction bias can be seen in Tables 2.16 and 2.19. The model that did not include finish time was not significant in predicting bias, $R^2 = .029$, $F_{(6, 298)} = 1.48$, $p = .184$, but the model that did include finish time was, $R^2 = .188$, $F_{(7, 297)} = 9.85$, $p < .001$. In correlational analyses and the regression model without finish time, no factor was a significant predictor of bias. Only gender showed a non-significant tendency for female runners to be more overconfident than male runners (Table 2.12). In contrast, numerous factors were significant contributors in the mode that accounted for finish time variance. High training volume and more months of running experience were associated with higher overconfidence or less underconfidence compared to low training volume and fewer months of running (Fig. 2.8A & 2.8B). Club members were consistently more overconfident than unaffiliated runners across finish times (Fig. 2.8C). Older runners showed a descriptive, but non-significant, tendency to be less overconfident or more underconfident than slower runners (Fig. 2.8D). Interestingly, though male runners were non-significantly less overconfident than female runners in the first model, they exhibited a non-significant tendency to be more overconfident or less underconfident after controlling for finish time in the second model (Fig. 2.8E). Finish time in itself was a significant predictor of bias, with slower runners being more likely to be overconfident than faster runners (Fig. 2.8F). Prediction lead was not a significant contributor to the model. Similar to the analyses for realistic prediction bias, these results illustrate the importance of accounting for finish time variance when examining the relationships between demographic factors and calibration.

Table 2.19

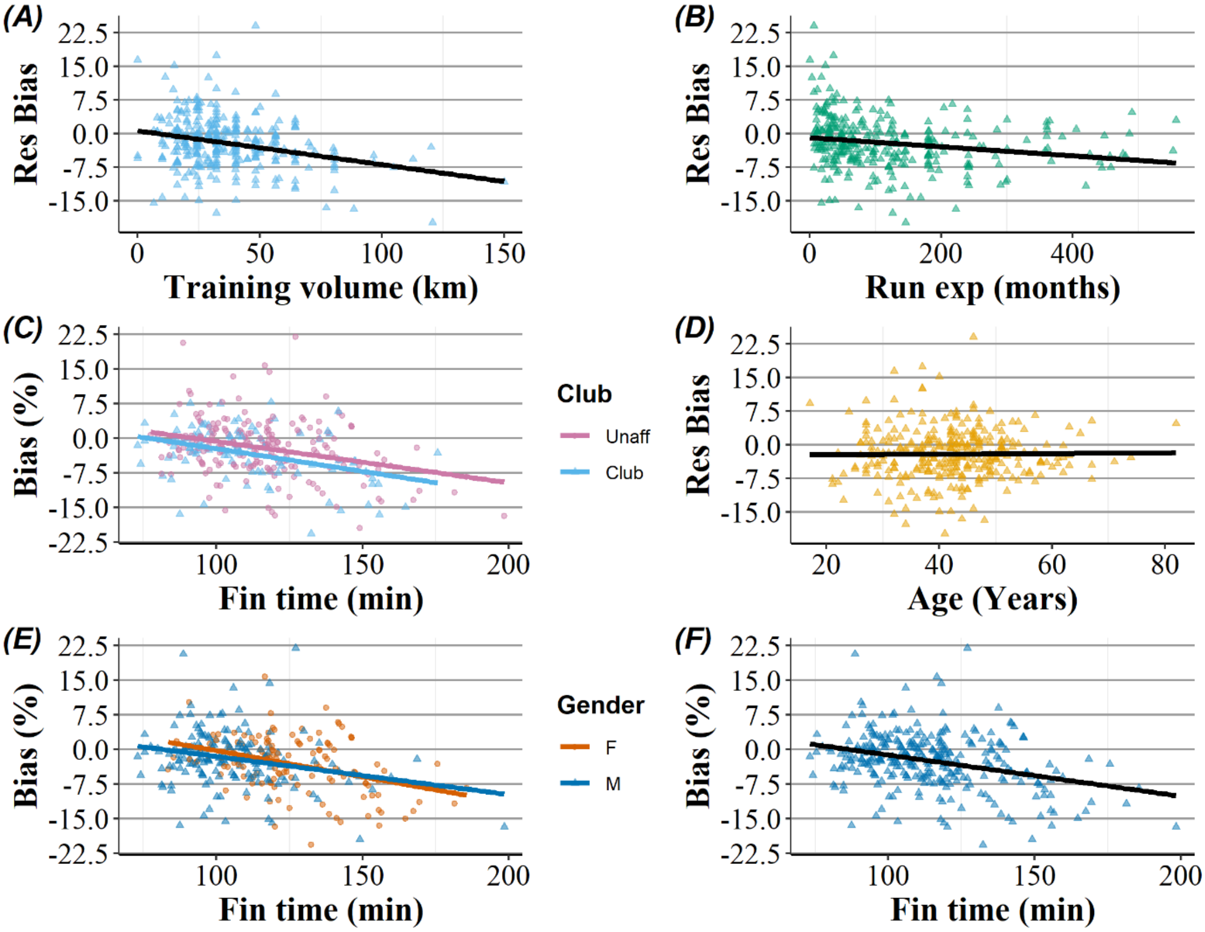
Multiple regression coefficients for demographic and prediction lead predictors on the outcome variable goal prediction bias.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No finish time	Intercept	-1.57	—	1.62	-0.97	.335
	km	0.00	-0.01	0.02	-0.18	.854
	Months	-0.01	-0.09	0.00	-1.48	.141
	Club	-1.09	-0.08	0.77	-1.41	.159
	Age	-0.01	-0.02	0.04	-0.26	.798
	Gender	1.24	0.10	0.72	1.73	.084
	PredLead	0.00	-0.03	0.01	-0.43	.669
Finish time included	Intercept	15.63	—	2.70	5.80	< .001
	km	-0.05	-0.18	0.02	-2.99	.003
	Months	-0.01	-0.18	0.00	-3.14	.002
	Club	-1.58	-0.12	0.71	-2.23	.026
	Age	0.06	0.11	0.04	1.82	.069
	Gender	-1.37	-0.11	0.74	-1.84	.066
	PredLead	0.00	-0.02	0.01	-0.44	.664
	FinTime	-0.14	-0.52	0.02	-7.64	< .001

Note. Demographic predictors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), PredLead to prediction lead, and FinTime to finish time.

Figure 2.8

Scatter plots illustrating the relationships between demographic factors and realistic prediction bias.



Note. Panel A shows the relationship between training volume and residual bias; panel B between months of running experience and residual bias; panel C between club membership and bias relative to finish time; panel D between age and residual bias; panel E between gender and bias relative to finish time; and panel F between finish time and bias. Residual bias across panels A, B, and D refers to bias minus the regression intercept and the finish time variance related to bias.

Absolute accuracy. Results from correlational analyses and the multiple regression model for goal prediction absolute accuracy can be seen in Tables 2.16 and 2.20. The model that did not include finish time was not significant in predicting absolute accuracy, R^2 = .026, $F_{(6, 298)}$ = 1.35, p = .234, but the model that did include finish time was, R^2 = .162, $F_{(7, 297)}$ = 8.21, p < .001. No factor, other than prediction lead, was significant in predicting absolute accuracy in the first model—these results were similar to the correlational analyses. However, numerous factors were significant predictors in the second model. After accounting for finish time variance in the model, age was a significant predictor, with older runners being more precise than younger runners (Fig. 2.9A). Male runners became significantly less precise than female runners, with this difference appearing to be driven by gender differences in faster, but not slower, finish times (Fig. 2.9B). Across models, runners were more precise in their goal predictions when they made them closer to the time of the race than when they made them earlier on (Fig. 2.9C). Finally, finish time was a significant predictor of absolute accuracy, with faster runners being more precise in their predictions than slower runners (Fig. 2.9D). No other factor had a significant contribution to goal prediction absolute accuracy.

Table 2.20

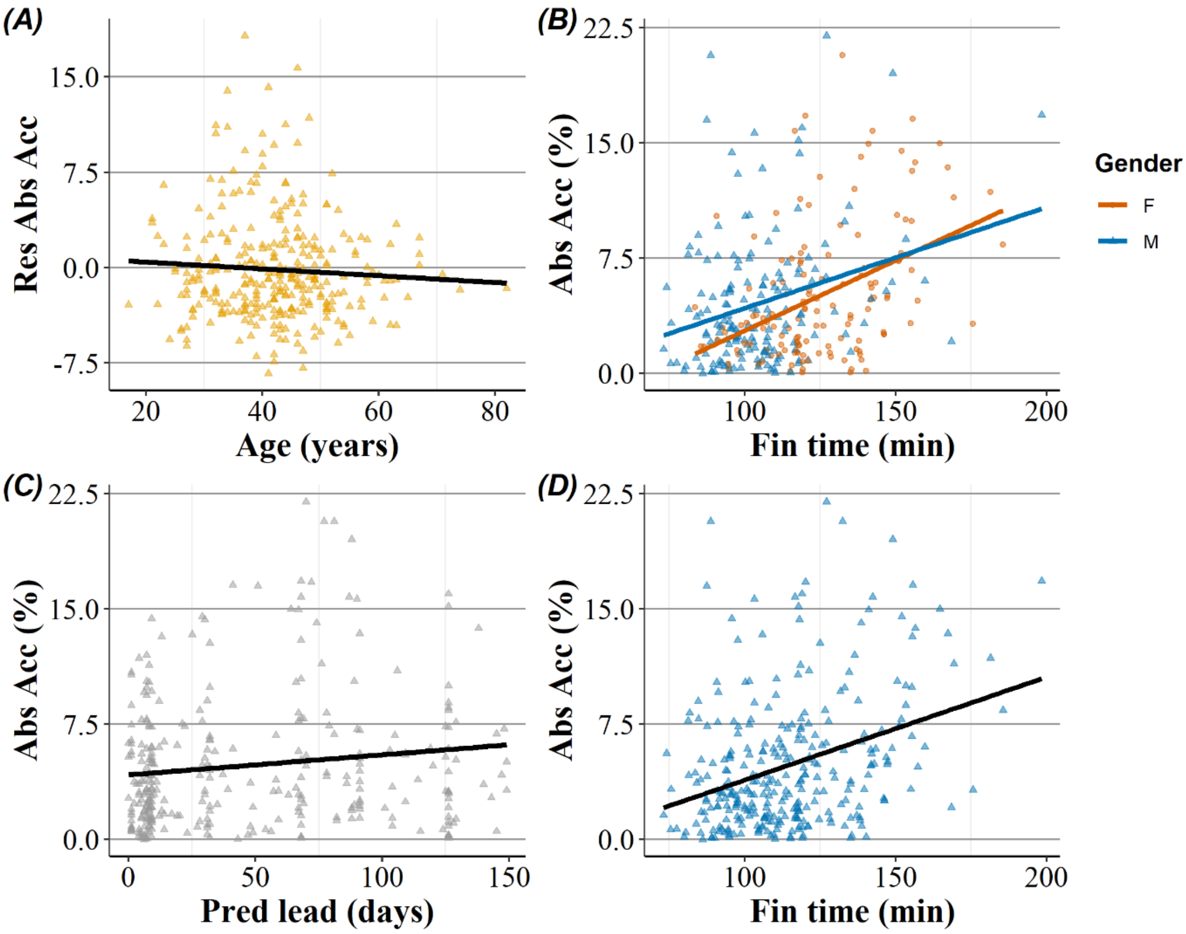
Multiple regression coefficients for demographic and prediction lead predictors on the outcome variable goal prediction absolute accuracy.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No finish time	Intercept	5.06	—	1.18	4.30	< .001
	km	-0.01	-0.07	0.01	-1.10	.274
	Months	0.00	-0.01	0.00	-0.21	.833
	Club	0.29	0.03	0.56	0.51	.610
	Age	-0.01	-0.01	0.03	-0.20	.844
	Gender	-0.29	-0.03	0.52	-0.55	.581
	PredLead	0.01	0.13	0.01	2.30	.022
Finish time included	Intercept	-6.44	—	1.99	-3.24	.001
	km	0.02	0.09	0.01	1.44	.151
	Months	0.00	0.07	0.00	1.19	.235
	Club	0.62	0.07	0.52	1.18	.239
	Age	-0.06	-0.13	0.03	-2.10	.036
	Gender	1.46	0.17	0.55	2.67	.008
	PredLead	0.01	0.13	0.01	2.44	.015
	FinTime	0.10	0.48	0.01	6.93	< .001

Note. Demographic predictors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. km is kilometres run per week, Months represents months of running experience, Club refers to club membership (Club = 0, if unaffiliated; Club = 1, if club member), Age to runner age, Gender to runner gender (Gender = 0, if female; Gender = 1, if male), PredLead to prediction lead, and FinTime to finish time.

Figure 2.9

Scatter plots illustrating the relationships between demographic factors and realistic prediction absolute accuracy.



Note. Panel A shows the relationship between training volume and residual absolute accuracy (i.e. absolute accuracy minus the regression intercept and the finish time variance related to absolute accuracy); panel B between gender and absolute accuracy relative to finish time; panel C between prediction lead and absolute accuracy; and panel D between finish time and absolute accuracy.

2.3.4 Discussion of Study 2

In Study 2, I aimed to test the reliability of findings from Study 1 by collecting demographic and prediction data online from runners who were participating in the Alloa Half Marathon. Additionally, I explored the extent to which prediction lead is associated with calibration.

As with Study 1, I anticipated that participants would produce overconfident goal predictions and less overconfident realistic predictions. Partially in line with my prediction, Goal predictions were overconfident, whereas realistic predictions were underconfident. Based on bias percentages, goal predictions were more overconfident (~2.6% relative to finish time) than realistic predictions were underconfident (~1.5% relative to finish time). Interestingly, precision was similar across prediction types, suggesting that the only difference between the two prediction types was in the direction of prediction bias.

The association between expertise and calibration I observed in Study 2 was in accordance with my expectations of faster runners being underconfident and precise, and slower runners being overconfident and less precise than faster runners. Faster runners were more precise than slower runners in their performance estimates across predictions. Furthermore, faster runners were underconfident and unbiased in their realistic and goal predictions respectively, whilst slow runners were overconfident across prediction types. Therefore, high expertise was associated with underconfidence or lack of bias and good precision, whereas low expertise was linked with overconfidence and poor precision.

My predictions that higher experience would predict better calibration than lower experience received mixed support. High training volume and long history of running experience were associated with overconfidence in realistic predictions, whereas low training volume and short running history were associated with underconfidence. When examining each factor individually (Table 2.16), high experience in terms of training volume and months of running was associated with higher realistic prediction precision. However, the contribution of training volume was no longer present after accounting for months of running experience and performance in the multiple regression models. The contribution of months of running experience on precision was in turn reduced after accounting for finish time in the second regression model. Training volume and months of running experience also contributed to goal prediction bias, but only after controlling for finish time variance. Higher experience predicted higher overconfidence across both factors. Neither factor predicted goal prediction absolute accuracy. Club membership was not a significant predictor of absolute accuracy for either prediction type, but, after controlling for finish time, club members were unbiased and overconfident compared to unaffiliated runners for realistic and goal predictions respectively. Experience findings suggest that all three factors were predictive of bias, regardless of prediction type, though finish time variance often had to be accounted for before these relationships became visible. Interestingly, training volume and months of running experience exhibited similar associations with calibration, though the latter factor was a stronger predictor of realistic prediction precision before accounting for performance.

For age, I expected that older runners would be better calibrated than younger runners, and I was interested in the extent to which this finding would remain present after controlling for other experience factors. After accounting for finish time, older runners were more likely to be underconfident or less overconfident than younger runners in their realistic predictions. Goal predictions showed a similar non-significant tendency. Age was also a significant predictor of goal, but not realistic, prediction absolute accuracy, with older runners being more precise than younger runners. These results were in line with my expectations, and suggested that, after controlling for performance, older runners are more likely to be underconfident or less overconfident and more precise than younger runners

I analysed gender differences in calibration to examine whether male participants would be more overconfident than female participants in a half marathon, which would be in line with marathon studies. I found that gender was an important predictor of calibration after accounting for performance variance. More specifically, male runners were less underconfident than female runners in their realistic predictions. This relationship appeared to be driven by faster, but not slower, finish times. Interestingly, female runners exhibited a non-significant tendency to be more overconfident than male runners for goal predictions when examined individually (Tables 2.12 & 2.16). However, the direction of this difference was reversed after accounting for finish time variance in the regression model, and male runners showed a non-significant tendency to be more overconfident than female runners instead. Gender did not contribute to absolute accuracy in realistic predictions, but it did in goal predictions. At faster finish times, male runners appeared to be less precise than female runners, whereas similar precision was exhibited for slower finish times. Overall, these results suggest that men are less likely to be underconfident and more likely to be overconfident than women when running performance variance is accounted for.

Finally, I expected that participants would be better calibrated the closer to the time of the race they made their predictions. Accordingly, prediction lead was a significant predictor of absolute accuracy for both prediction types, but it did not have a clear influence on bias. Runners who made their predictions closer to the time of the race were more precise than runners who made their predictions further ahead in advance. It is thus possible that runners have more information about their physical capacity and race conditions closer to the time of the race, allowing them to make better informed, and thus more accurate, performance estimates.

2.4 GENERAL DISCUSSION

The present chapter aimed to improve our understanding of the factors that contribute to running calibration. I presented and analysed data from two running events of different distances: the 2018 Edinburgh Christmas 10k Run and the 2019 Alloa Half Marathon. I anticipated that high expertise and experience would be associated with better calibration than low expertise and experience. I was also interested in the extent to which older age would predict better or poorer calibration after accounting for other experience factors. Given previous gender findings on pacing and bias, I wanted to explore whether male runners would be more overconfident than female runners, replicating marathons findings, or whether this relationship would be weakened or not be present for shorter distances. In Study 2, I also examined the influence of prediction lead on calibration, expecting that predictions made closer to the time of the race would be more accurate than predictions made earlier in advance. Finally, I anticipated that runners would make overconfident goal predictions, and less overconfident realistic predictions across studies.

Prediction type affected calibration in both studies. In Study 1, participants were underconfident by ~3.9% of their performance in their realistic predictions, and unbiased in their goal predictions (~0.4%). Furthermore, goal predictions were more precise than realistic predictions. Conversely, in Study 2, participants were overconfident (~-2.6%) in their goal predictions, and, though they were underconfident in their realistic predictions, the magnitude of this underconfidence was smaller than in Study 1 (1.5% vs 3.9%). Absolute accuracy was similar across prediction types. The finding that realistic predictions in both studies were underconfident was surprising, and contrasted previous running results where participants received instructions to be as accurate as possible and were still overconfident (Krawczyk & Wilamowski, 2016, 2018). Though it is not clear why participants in the two studies made underconfident realistic predictions, it is possible that asking participants to make two predictions, rather than one, affected the way they approached their predictions. Since races are competitive events for which runners train and set goals (especially for longer courses, such a marathons), when asked for a single prediction, runners might be more likely to make an estimate that closely resembles their performance goals. By providing them with the opportunity to make both goal and realistic predictions instead, we can control for this possibility, and thus reduce realistic prediction overconfidence. Additionally, instructions for goal predictions were presented before realistic predictions and, despite asking runners to consider both before making their estimates, they might have decided on a goal prediction first, which they then just adjusted for their realistic prediction. Overly high adjustments for realistic predictions could explain the observed underconfidence.

Interestingly, runners in Study 1 were not overconfident in their goal predictions, whereas runners in Study 2 were. This can possibly be explained by the 2018 Edinburgh Christmas 10k Run being significantly faster than the 2017 and 2019 races, suggesting better-than-usual performance for the course. Runners might have thus performed better than they would have expected given the race’s typical conditions, which would explain the lack of overconfidence in goal predictions and the high underconfidence in realistic predictions. In contrast, performance in the 2019 Alloa Half Marathon was not atypical of the course, as it was similar to 2016 and 2017. Consequently, differences in the extent to which race conditions were typical or atypical of each course could have contributed to calibration differences between the two races. Overall, prediction type results from the two races indicate that the extent to which athletes opt to make realistic or goal predictions for their performance can have a substantial effect on their calibration. Nonetheless, this is the first examination of prediction type’s contribution to calibration, and thus the relationship warrants further investigation. Future studies need to collect data from more races, control for order of instructions (i.e. by counterbalancing), and compare how runners make predictions when asked to only make one prediction type compared to when they have to make both.

In accordance with my predictions and previous cognitive and exercise research (e.g. Dunning et al., 2003; Kolovelonis, 2019; Kolovelonis & Goudas, 2018; Kruger & Dunning, 1999; Schlösser et al., 2013), high expertise was associated with underconfidence (or lack of bias) and high precision, whereas low expertise was associated with overconfidence (or less underconfidence) and low precision. In Study 1, slower runners made less underconfident realistic predictions than faster runners, though there was no finish time relationship with absolute accuracy. For goal predictions, faster runners were more likely to be underconfident and precise compared to slower runners, who tended to be overconfident and less precise. In Study 2, faster runners were more precise than slower runners for both prediction types. Additionally, faster runners were underconfident and unbiased for realistic and goal predictions respectively, whilst slower runners were overconfident for both prediction types. The present studies were thus the first to provide strong and clear support for the relationship between expertise and calibration in running, indicating that high expertise is linked with positive calibration outcomes.

I expected experienced runners who ran more kilometres per week to be better calibrated than less experienced runners. In Study 1, higher training volume predicted higher prediction precision than lower training volume across prediction types. However, this association became non-significant after controlling for finish time, suggesting that expertise accounted for the training volume influence on precision. Training volume did not predict goal prediction bias, and high training volume only showed a descriptive, but non-significant, association with lower realistic prediction underconfidence after accounting for finish time variance. In Study 2, high and low training volumes were associated with overconfidence and underconfidence respectively across prediction types. Training volume only predicted absolute accuracy for realistic predictions, and this relationship was not significant after accounting for the variance of months of running experience. Overall, though training volume can contribute to running calibration, the magnitude of this relationship appears to be small (precision) and inconsistent (bias findings were limited to Study 2). Consequently, we need to also account for other experience and expertise factors when we examine running calibration.

The relationship between months of running experience and calibration was not consistent across the two studies. In Study 1, months of running experience did not predict bias or absolute accuracy for either prediction type. In Study 2, runners who had been running for more months were more likely to be overconfident for both prediction types compared to runners who had been running for fewer months. Additionally, though length of running experience did not predict absolute accuracy for goal predictions, it did for realistic predictions. More experienced runners were more precise than less experienced runners, even after accounting for the variance of other experience factors. However, the inclusion of finish time in the model diminished this relationship, suggesting that, as with training volume in Study 1, expertise accounts for the contribution of experience factors to precision. The reason behind the results discrepancy between the two studies is unclear. It is possible that because runners completed this information right before participating in the race in Study 1, they were less likely to remember it accurately, whereas they had time to verify it when they completed it online. Overall, months of running experience had a similar relationship with calibration as training volume did in that, though it can inform us about certain calibration structures, its contribution is relatively minor.

Club membership played a role in calibration in Study 2, but not Study 1. For the Edinburgh Christmas 10k Run, club members exhibited similar bias and precision as unaffiliated runners across prediction types. For the Alloa Half Marathon, club membership also failed to predict absolute accuracy. However, club members were unbiased and overconfident for realistic and goal predictions respectively, whereas unaffiliated runners were underconfident and less overconfident. The present findings did not support Liverakos and colleagues' (2018) observation of club members being more precise than unaffiliated runners. This could result from club membership being a secondary experience factor whose contributions to performance and calibration are accounted for by training volume and months of running experience. Low power in the present studies could also explain this discrepancy, as a power analysis using the effect size reported by Liverakos and colleagues suggested that, for a power of 80%, I would have needed a sample of at least 310 club members and 310 unaffiliated runners. The present samples only consisted of 38 and 97 club members respectively. Therefore, a much larger sample size is required before we make strong conclusion about the role of club membership in running calibration.

Overall, experience contributions to precision appear to be small and can often be accounted for by expertise measures. It is thus important not to rely on experience alone when assessing exercise calibration. This conclusion is in line with previous research on exercise calibration and pacing, where individual markers of experience only had a relatively minor impact (i.e. small effect sizes) on the outcome variables (Deaner et al., 2014; Kolovelonis, 2019; Liverakos et al., 2018; Swain et al., 2019). An interesting and unexpected finding from Study 2 was that experienced runners were more likely to be overconfident (or less likely to be underconfident) than less experienced runners. The extent to which this is a reliable result needs to be further tested, as it can shed more light on how experience markers contribute to calibration.

For age, I anticipated that older runners would be better calibrated than younger runners, though I was uncertain as to whether this relationship would remain present after controlling for other experience factors. In Study 1, though age did not contribute to realistic prediction bias, older runners were more underconfident (or less overconfident) in their goal predictions than younger runners. In the same vein, after controlling for finish time, older runners in Study 2 were more likely to be underconfident (or less overconfident) in their realistic predictions than younger runners, and showed a similar descriptive, but non-significant, tendency for goal predictions. Age did not predict realistic prediction absolute accuracy in either study. However, older runners were more precise in their goal predictions than younger runners across studies. These findings support previous results of age's positive influence on calibration and pacing (e.g. Deaner et al., 2014; Liverakos et al., 2018; March et al., 2011; Trubee et al., 2014), as well as the suggestion that age can make a positive contribution to calibration in naturalistic settings (Cauvin et al., 2019; Devolder et al., 1990). They are also not consistent with cognitive evidence of increased overconfidence in older adults (e.g. Cauvin et al., 2019; Soderstrom et al., 2012). Interestingly, age's influence on calibration was present even after controlling for other markers of experience. It is thus possible that age can serve as an independent contributor to calibration, regardless of years of experience. Facing physical decline, older athletes might have to rethink their training and performance capacity in a way that enhances their metacognitive awareness. Adjustments for physical decline could also explain why older adults were more likely to be underconfident or less overconfident than younger adults.

Results from both studies generally exhibited patterns of relative male overconfidence, but only after accounting for finish time variance. In Study 1, male runners were less underconfident than female runner in their realistic predictions. This appeared to be driven by slower finish times, where women were underconfident and men were slightly overconfident. Runners of either gender were similarly underconfident for faster finish times. In Study 2, men were less underconfident in their realistic predictions than women for faster finish times, but similarly biased for slower finish times. For goal predictions, male runners in Study 1 were less likely to exhibit underconfidence than female runners. This relationship seemed to be driven by gender differences in faster, but not slower, finish times. In Study 2, women exhibited a non-significant tendency to be less overconfident in their goal predictions than men. Gender did not contribute to realistic prediction absolute accuracy for either study, but it did for goal predictions. In both studies, men were less precise than women for faster finish times, but similarly precise for slower performance.

These findings suggest that male runners were less likely to be underconfident and more likely to be overconfident than female runners across different race lengths, supporting previous running calibration and pacing literature (Deaner et al., 2016, 2014; Deaner & Lowen, 2016; Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018; March et al., 2011; Trubee et al., 2014). Though calibration studies have typically found consistent gender differences across finish times (Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018), the present results highlight that, when we examine gender differences in calibration, we need to also account for performance. This is in accordance with pacing research where gender differences in bias might depend on finish time (e.g. Deaner et al., 2016, 2014; Deaner & Lowen, 2016). However, as I discussed more extensively in Section 1.3.1.4, Deaner and colleagues (2014) have suggested that we need to adjust women's finish times by a theoretical value of 12% to fully account for gender differences in performance capacity. Otherwise, they argued that gender differences in bias and pace slowing are likely to be exaggerated. Since I did not adjust women's finish times in the present studies (for justification, see the footnote in Section 1.3.1.4), I acknowledge that it is possible for the results to have overestimated the magnitude of gender differences in bias for 10km and half marathon races. Nonetheless, pacing studies have generally found gender differences in pace slowing to be present even after adjusting women's finish times (e.g. Deaner et al., 2016, 2014; Deaner & Lowen, 2016), so gender differences in bias and pace slowing are likely to be reliable.

The result that male runners were less precise for faster finish times for goal, but not realistic, predictions is novel, as this relationship has not been examined in exercise before. It would be interesting to further test its reliability to confirm whether male runners are actually less precise when they set goals than female runners. Overall, the present studies support relative male overconfidence resulting from lower underconfidence or higher overconfidence for male athletes compared to female athletes across a range of race lengths. Since gender differences in glycogen depletion (see Section 1.3.1.4) are only able to explain male overconfidence in longer (e.g. marathons), but not shorter (e.g. 10km and half marathons), distances, it is likely that higher male overconfidence and pace slowing also result from psychological (e.g. risk tasking or competitiveness) differences between the two genders.

I only examined the relationship between prediction lead and calibration in Study 2. Previous running calibration studies have either collected predictions right before the event examined (Krawczyk & Wilamowski, 2016, 2018), or have not had information regarding the time when participants made their predictions (Hubble & Zhao, 2016; Liverakos et al., 2018). Because of this, the extent to which runners might be more accurate when they make predictions closer to the time of the race has not been investigated before. Study 2 showed that, though prediction lead did not exhibit a significant association with bias, runners who made realistic and goal predictions closer to the time of the race were more precise than runners who made their predictions earlier in advance. This means that runners who make their predictions closer to the time of the race are more likely to make precise performance predictions than runners who make theirs earlier in advance. This finding has important implications for athletes and event organisers. As they get closer to an event, athletes should constantly use training information to adapt their predictions and strategies in order to maximise accuracy and strategy effectiveness. Similarly, event organisers should take in consideration prediction lead when allocating runners in starting placements to ensure optimal accuracy in their decisions. Therefore, when performance predictions are not made within near time proximity to a competition, information about prediction lead should also be collected and used for accuracy evaluation.

The present chapter had important implications for understanding running calibration, as I was able to examine how prediction type, demographic factors, and prediction lead contribute to prediction accuracy. The extent to which runners rely on realistic or goal predictions can have a considerable impact on their strategic planning for both training and competitions, thereby affecting performance, motivation, and injury risk. Since this relationship has not been examined before, we need to conduct more studies that further assess how different prediction types influence calibration by expanding on the methods used in the present studies. Furthermore, present results highlight the importance of specifying prediction type when examining calibration to ensure that runners do not opt for different prediction types that could introduce confounds to the data. Similarly, we need to account for prediction lead when we collect data from competitive events, as the time when athletes make their predictions can influence precision.

The studies presented were useful in identifying the influence of demographic factors on calibration. Expertise was a strong predictor of calibration, and often accounted for the influence of other experience factors on precision. Additionally, controlling for performance was essential to observe the relationships between certain variables (e.g. gender) and calibration. Expertise accounting for experience associations with precision does not necessitate that experience is not an important contributor to calibration. Instead, it supports the argument that we need to use numerous markers of experience in conjunction with other variables to predict calibration. Furthermore, we can assume that all participants in the studies had substantial running experience, as they were taking part in running competitions, rendering it difficult to strongly dissociate between experience and expertise. It would thus be interesting to examine whether the influence of experience on calibration is stronger when examining athletes with similar performance capacity, but very different experience levels. Additionally, age was an independent contributor to calibration, even after accounting for other experience factors. This could suggest that older runners have to increase their metacognitive performance awareness to cope with physical decline. Finally, male runners were more overconfident or less underconfident than female runners, suggesting that athletes, fitness instructors, event organisers, and coaches, among others, need to account for gender differences in bias when it comes to self-regulation. It is important to ensure that female and male athletes avoid setting underconfident and overconfident goals respectively, as these can lead to ineffective strategies and suboptimal performance (see Section 1.2.3).

2.5 CONCLUSION

Findings from the present chapter were important in understanding how prediction type, demographic factors, and prediction lead influence running calibration, and can thus be used to improve exercise performance and motivation, and reduce injury risk. It should be noted that, since each factor examined might have limited impact on calibration when examined individually, it is essential to identify and use as many factors as possible to produce more detailed accounts on how athletes assess their performance. Furthermore, Chapter 2 only focused on calibration in running. As illustrated in Chapter 1, calibration in other exercise modalities does not always show the same relationships with demographic factors as in running. Because of this, we need to also examine the relationships between demographic factors and calibration in sports other than running. This will allow us to complement the present findings and to better understand how each demographic factor contributes to calibration across different types of physical activity. This is the focus of Chapter 3.

CHAPTER 3: DEMOGRAPHIC FACTORS & HIGH-INTENSITY FUNCTIONAL MOVEMENT EXERCISE CALIBRATION

3.1 INTRODUCTION & RATIONALE

In Chapter 2, I examined the relationship between demographic factors and running calibration. Results were important in identifying variables that we can use to predict whether runners will exhibit good or poor calibration. Given the range of evidence and suggestions from Chapter 1 on calibration being an important contributor to performance, motivation, and injury risk (see Section 1.2.3), then knowing which factors we can use to assess calibration is essential in understanding how to produce optimal running outcomes. For example, based on expertise findings from Chapter 2, we can expect faster runners to be well calibrated and more likely to implement appropriate training and competition strategies that facilitate performance. Conversely, we would expect slower runners to be overconfident and imprecise in their performance estimates, rendering it difficult for them to implement effective strategies, thus limiting the extent to which they can self-optimise their performance. Though Chapter 2 focused on running, the benefits of good calibration are not restricted to one exercise modality—or the exercise domain altogether. We thus need to be in a position to evaluate calibration across exercise modalities.

A way of assessing calibration across exercise modalities would be to use findings from one modality to make inferences about another. In this case, we would use calibration associations with demographic factors identified in running to inform us on the role of the same factors in calibration for tasks and skills in other physical activities, e.g. basketball, tennis, and golf.[3] However, the extent to which calibration results from one exercise modality apply to another is not clear. In Section 1.3.1, I noted that, on certain occasions, there were differences between running and other physical activities (e.g. basketball, tennis, and golf) in the relationships between demographic factors and calibration. For example, in calibration and pacing studies, there was a clear tendency for female runners to exhibit lower overconfidence or pace slowing than male runners (Deaner et al., 2016, 2014; Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018; March et al., 2011). In contrast, physical education studies using basketball shooting and dribbling tasks did not exhibit gender

[3] As mentioned in Section 1.3.1.2, when I refer to sports such as tennis and basketball, I only refer to skills and tasks that do not involve predicting and assessing opponent and teammate behaviour. Calibration in tasks where performance is determined by the actions of others is outside the scope of the present thesis.

differences in bias (Kolovelonis, 2019; Kolovelonis & Goudas, 2018; Kolovelonis, Goudas, & Dermitzaki, 2012). Similarly, though high expertise was consistently associated with less pace slowing in running than low expertise (Breen et al., 2018; March et al., 2011; Nikolaidis & Knechtle, 2017, 2018b), this relationship was inconsistent for tasks and skills in modalities such as tennis and golf (Fogarty & Else, 2005; Fogarty & Ross, 2007). It is thus important to investigate why such differences in calibration findings have been observed between exercise modalities.

One explanation for the discrepancies presented above is that research on the relationships between demographic factors and exercise calibration is scarce, and thus not sufficient in informing us about its generalisability across exercise modalities. To address this, we need to conduct more studies that examine how demographic factors influence calibration in different athletic tasks. Another explanation is that differences in samples used could contribute to calibration discrepancies between modalities. For example, running studies typically recruit adults of a wide range of ages (e.g. Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018; Liverakos et al., 2018), whereas physical education studies tend to recruit younger participants (e.g. children aged between 10-12 years; e.g. Kolovelonis, 2019; Kolovelonis & Goudas, 2018; Kolovelonis, Goudas, & Dermitzaki, 2012). If any gender differences in sports bias arise during adolescence or early adulthood, then this will explain why higher male overconfidence is observed in the former, but not the latter.

Inconsistencies in the measures used could also contribute to discrepancies between studies. In this vein, Krawczyk and Wilamowski (2016) examined expertise influence on running bias by using 21km split times in marathons and found that faster runners were less overconfident than slower runners. Similarly, pacing studies in running have often used finish times to investigate the association of expertise with pace slowing and have typically found faster runners to exhibit less pace slowing than slower runners (e.g. Deaner et al., 2014; March et al., 2011). In contrast, Fogarty and Else (2005) operationalised golf expertise in terms of participant golf handicap (e.g. lower handicap indicated higher expertise), and Fogarty and Ross (2007) divided participants into groups of experts and non-experts based on whether they were current or former professional players (experts), or junior or social players (non-experts). Neither study found a clear influence of expertise on calibration, as expertise was not associated with calibration in golf, and expert tennis players were better calibrated than non-experts in the difficult, and not the easy, tennis serving task used. Given the discrepancies in calibration results between studies using performance and studies using other measures to operationalise expertise, it is possible that the way in which we assess expertise contributes to whether we observe a relationship between expertise and exercise calibration.

Exercise complexity is also a potential reason for inconsistencies in calibration findings between different exercise modalities. Though running is a demanding activity, it is relatively simple to perform in comparison to complex activities that require both high physical fitness, and motor coordination (e.g. basketball, tennis, and golf)—even in the absence of competitors. High complexity could render performance in an activity difficult to predict, because athletes have to consider more factors than they do in simpler activities. Differences in complexity could thus make it easier for athletes to track and predict their running rather than their basketball shooting or tennis serving performance. If this is true, then exercise modality could affect the relationship between demographic factors and calibration. Unfortunately, as discussed in the previous paragraph, the scarcity of exercise calibration research does not allow us to reach a strong conclusion regarding the role of activity complexity in the discrepant findings observed. It is thus important to conduct studies that also explore the relationship between demographic factors and calibration in complex and unpredictable tasks. To achieve this, I collected and analysed calibration data from high-intensity functional movement exercise (HIFME), which I present here.

HIFME is a popular form of exercise that consists of various exercise modalities, such as CrossFit, body sculpt, and circuit training. HIFME is not a widely used term, but rather an inclusive one that I have devised to encapsulate a range of similar exercise regimes. It can be defined as high-intensity exercise that takes place during a short period of time (typically less than an hour) and consists of gymnastics, athletics, and weightlifting (Butcher, Neyedly, Horvey, & Benko, 2015; Claudino et al., 2018; Gianzina & Kassotaki, 2019; Paine, Uptgraft, & Wylie, 2010; Weisenthal, Beck, Maloney, DeHaven, & Giordano, 2014). HIFME thus combines aerobic and anaerobic exercise elements, creating varied and complex workouts. HIFME workouts can also be very unpredictable, as the combination of a wide range of movements allows each session to have a unique structure. Accordingly, HIFME typically aims to increase functional fitness, which refers to the adaptation to different exercise situations and requirements. For the above reasons, I examined how demographic factors influence calibration in complex and unpredictable workouts using HIFME.

Though HIFME's complexity and unpredictability renders it a salient exercise modality to examine the generalisability of the associations between demographic factors and calibration across different sports, there are more reasons to examine HIFME calibration, which I outline below. HIFME as an exercise modality was introduced to the fitness world relatively recently, and has been continuously growing in popularity. In 2014, 200,000 athletes signed up for the CrossFit Open (a competition where athletes submit their workouts online), and there were 11,000 active CrossFit affiliates in the world (Butcher et al., 2015). In 2016, the number of affiliates rose to 13,000 (Meyer, Morrison, & Zuniga, 2017), whilst over 300,000 athletes participated in the CrossFit Open in 2019 (Henderson, 2019). It is thus important to identify the factors that influence HIFME calibration, and use them to optimise athlete training and performance. Additionally, HIFME's combination of numerous challenging exercises and high intensity can contribute to elevated injury risk. Though reviews have generally shown HIFME to have similar injury rates as some other exercise modalities, e.g. gymnastics and powerlifting, concerns about injury risk remain valid (Gianzina & Kassotaki, 2019; Meyer et al., 2017; Weisenthal et al., 2014). It is thus important for athletes to understand their limits and exercise at appropriate intensities. In Section 1.2.3.6, I discussed the implications of calibration for injury risk in sports, arguing that good calibration is essential for the selection of appropriate training and competition loads, and thus reduced injury risk. Well-calibrated HIFME athletes should thus be able to both optimise their performance and reduce their susceptibility to injuries.

3.2. STUDY 3 – DEMOGRAPHIC FACTORS & HIFME CALIBRATION

3.2.1 Study specifics

The aim of Study 3 was to expand previous literature on the relationships between demographic factors and calibration by analysing data collected from a complex exercise modality. To achieve this, I examined the influence of expertise, experience, and gender on HIFME calibration. To assess experience, I recruited participants who had engaged in HIFME before (for at least a month), and participants who had not participated in HIFME before, but still engaged in physical exercise. Within the group with a HIFME background, I also collected data on months of HIFME experience, and number of HIFME sessions completed per week. Based on previous running and physical education research, I anticipated that participants with high expertise (i.e. better performance) and participants in the HIFME group would be better calibrated than participants with low expertise (i.e. poorer performance) and participants in the non-HIFME group respectively. Additionally, I expected that more experienced participants within the HIFME group would be better calibrated than less experienced participants. For gender, I was interested in whether male participants would be more overconfident than female participants—in line with running findings—or whether they would be similarly biased—in line with physical education findings. To control for prediction type, I instructed all participants to provide realistic, and not goal, predictions. Furthermore, participants made predictions for and completed two HIFME workouts with different structures to examine calibration patterns in different types of workouts within the same exercise modality. Results from the present study had important implications for understanding how demographic factors are associated with calibration in a complex and unpredictable exercise modality, and whether these associations reflect findings from other modalities. Finally, examining calibration in HIFME is important in assisting athletes who engage in this novel exercise modality with optimising performance, and reducing injury risk.

3.2.2 Methods

3.2.2.1 Participants

Sixty participants between the ages of 18 and 40 years old took part in Study 3 (29 men and 31 women; 30 with and 30 without HIFME experience; M_{age} 24.6 years old, *SD* = 5.0 years). I placed participants in the HIFME experience group if they had previously engaged in exercise that combined aerobic and anaerobic training at the same time during a workout (e.g. CrossFit, body sculpt, and circuit training) consistently and for at least a month (e.g. for more than a total of ten hours). Participants in the non-HIFME group needed to have had experience with aerobic and anaerobic exercise, but not at the same time (e.g. engage in running and weightlifting without combining them in workouts). I recruited all participants from the University of St Andrews and local population. As part of Study 3, I asked participants to complete two workouts, hereon referred to as the AMRAP (As Many Repetitions as Possible) and the Rounds workouts (see Section 3.2.2.2 for details).

Of the sixty participants recruited, fifty-five completed the AMRAP workout, fifty-seven the Rounds workout, and fifty-three both of them. Participants who did not complete a workout either stopped because of fatigue, or because they could not achieve a movement standard (e.g. squat depth; see Section 3.2.2.2 for movement standards). I classified participants as outliers if the precision of their predictions was very poor (i.e. *z* scores of absolute accuracy percentages relative to performance were higher than three). I removed two outliers (two men who did not have HIFME experience) from data analysis for the AMRAP workout, leading to a sample of fifty-three participants (25 men and 28 women; 29 with and 24 without HIFME experience, M_{age} 24.7 years old, SD = 5.0 years). I removed one outlier (a woman without HIFME experience) from the Rounds workout, leading to a sample of fifty-fix participants (29 men and 27 women; 30 with and 26 without HIFME experience, M_{age} = 24.8 years old, SD = 5.1 years). Overall, I removed three outliers (two men and one woman, all without HIFME experience) from both workouts, leading to a sample of fifty participants (25 men and 25 women; 29 with and 21 without HIFME experience; M_{age} 24.8 years old, SD = 5.0 years) for calibration comparisons between the two workouts.

This study received approval by the University of St Andrews School of Psychology & Neuroscience Ethics Committee (Ethics approval code: PS13328; see Appendix 8.1.3). Participants were compensated at a rate of £5/hour.

3.2.2.2 HIFME workouts

As Many Repetitions As Possible (AMRAP). In the first workout, participants had to complete as many rounds of 5 inverted rows (Fig. 3.1), 10 burpees (Fig. 3.2), and 15 air squats (Fig. 3.3) as they could within 10 minutes. The standards for each movement were the following: for inverted rows, participants had to fully extend their arms at the bottom of the position and then get their chest as close to the bar as possible; for burpees, participants had to touch the floor with their chest, stand up, and then jump and clap at the top of the position; for air squats, participants had to squat deep enough for the crease of the hips to be lower than (or at least at the same height as) the top of the knee. Failure to complete a repetition in accordance with the above standards led to its repetition. I measured predictions and performance in terms of total repetitions completed. When participants predicted or completed partial rounds, I added the repetitions completed to the sum of repetitions from the completed rounds.

Figure 3.1

Demonstration of inverted rows.

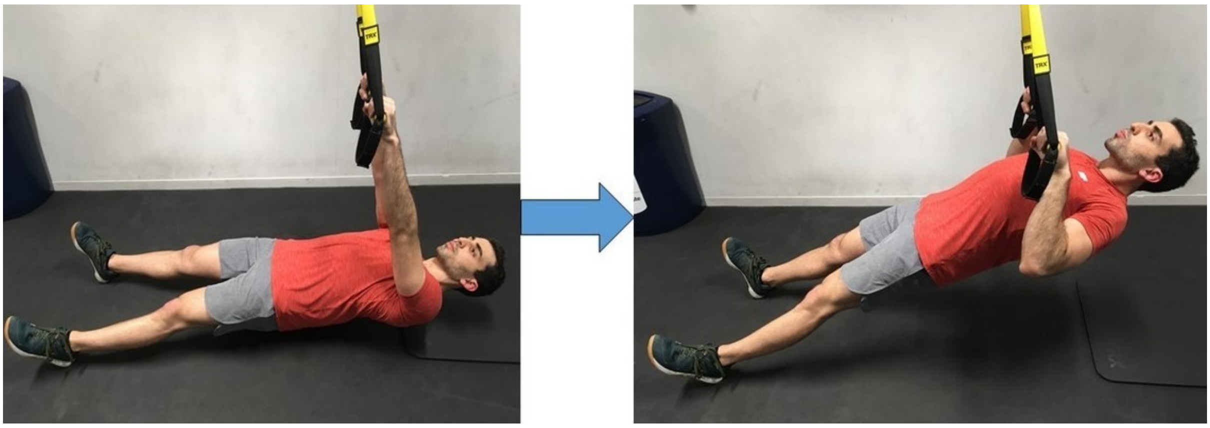


Figure 3.2

Demonstration of burpees.

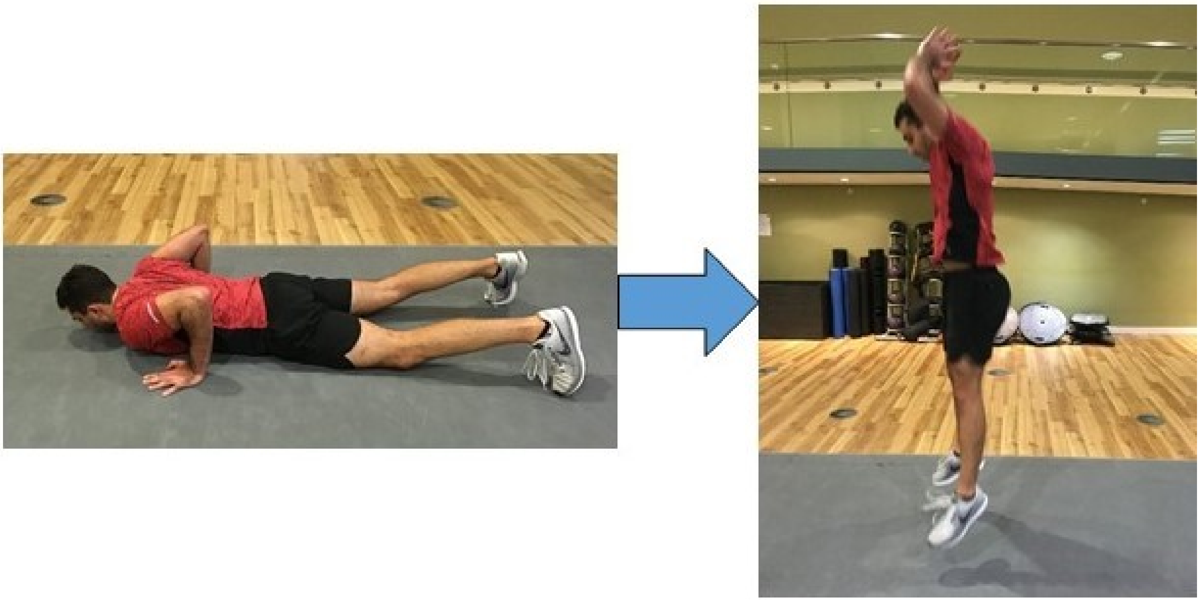


Figure 3.3

Demonstration of air squats.

Five rounds. Participants had to complete five rounds of 5 push-ups (Fig. 3.4), 10 calories on a rowing ergometer (typically fewer than 20 rows to complete), and 15 sit-ups (Fig. 3.5) as fast as possible. The standards for each movement were the following: for push-ups, the chest had to be as close to the floor as possible (this depended on each participant's strength and mobility); for rowing, there was no movement standard and participants just had to complete the number of calories required (they were familiarised with this unit of measurement during a warm-up session); for sit-ups, participants had to touch the floor with their upper back on the way down, and their feet with their hands at the top position. Failure to complete a repetition in accordance with these movement standards led to its repetition. I measured predictions and performance in terms of time (in seconds) taken to complete all five rounds. If a participant failed to complete all five rounds within 15 minutes, I did not register a score.

Figure 3.4

Demonstration of push-ups.

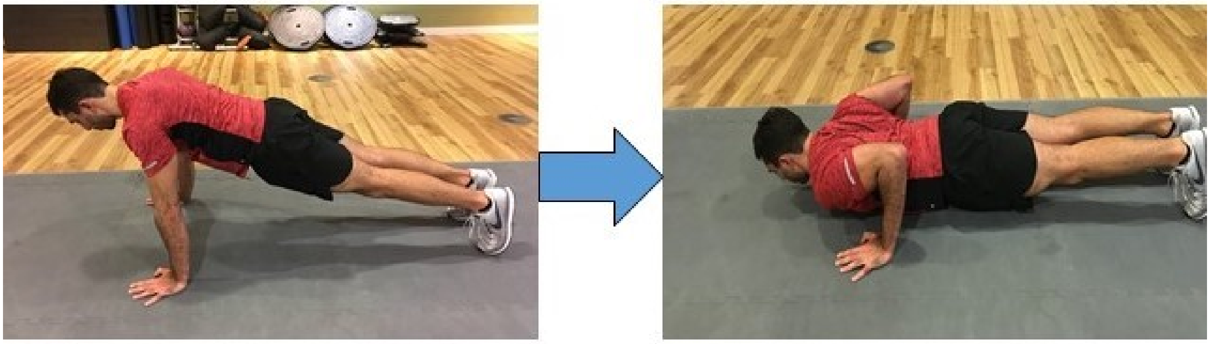
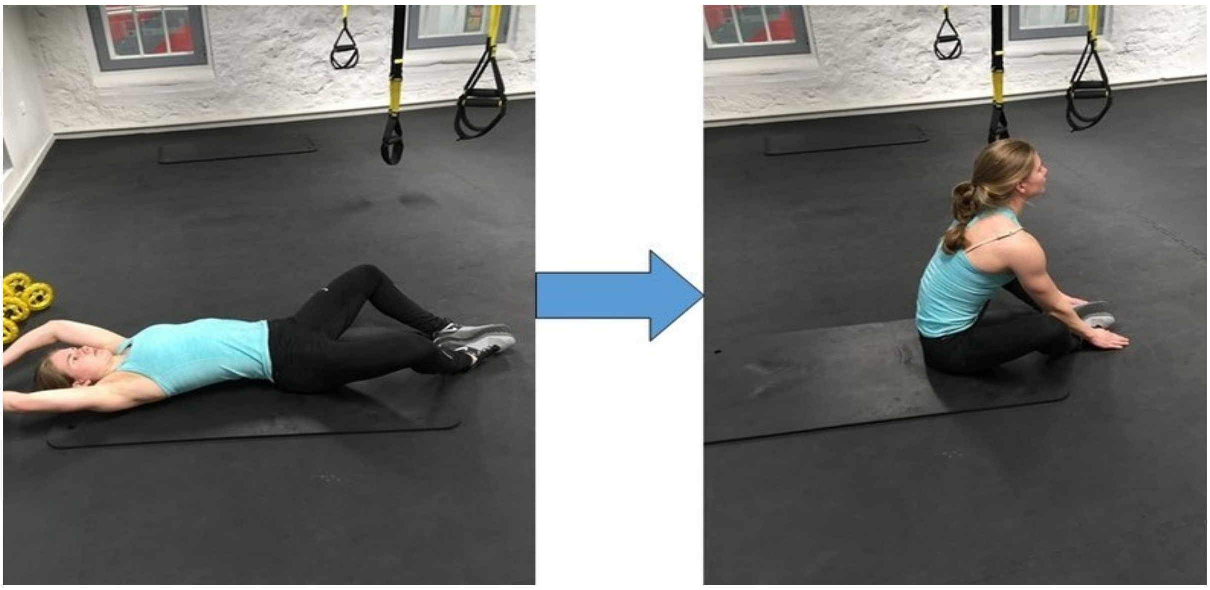


Figure 3.5

Demonstration of sit-ups.



Each workout had a different structure, but I intended for both to be of similar difficulty. I used two workouts with different formats because of the very variable workout format typically implemented in HIFME. In line with the HIFME definition used in the present thesis, both tasks combined aerobic and anaerobic exercise elements, whilst targeting numerous muscle areas. Participants had a five-minute resting period between workouts. I counterbalanced the order of workouts to limit the influence of order and fatigue effects on calibration.

3.2.2.3 Materials

Rowing in the Rounds workout was conducted on a Concept 2 model D rowing ergometer, used for indoor rowing ("Model D," n.d.). The resistance was set at 7/10 (common resistance used in CrossFit workouts). A 20kg barbell and a squat rack, or a pair of TRX if the other equipment was not available, were used for 'inverted rows.' A small exercise pillow was used for sit-ups in the Rounds workout to ensure participants did not experience lower back discomfort.

I collected ratings of perceived exertion before and after each workout using the Borg scale, which incorporates values ranging from 6 (representing very, very light exertion; e.g. lying in bed) to 20 (representing very, very hard exertion; e.g. final sprint of a long race; Borg, 1982; see Appendix 8.4).

3.2.2.4 Design

The focus of the present analysis was on the influence of expertise, experience, and gender on calibration bias and absolute accuracy. I operationalised expertise in terms of performance (i.e. number of repetitions completed for the AMRAP workout, and finish time for the Rounds workout). I assessed experience based on participants' self-reports of exercise background. I placed participants who reported previous experience with exercise modalities that satisfied the HIFME criteria presented in Section 3.2.2.1 in the HIFME group, and participants who did not report such experience in the non-HIFME group. I also used months of HIFME experience and number of HIFME training sessions per week as experience factors in the HIFME group alone. I did not examine age influence on calibration in the present sample, as all participants were within a narrow age range (18-40 years old). As a secondary analysis, I explored whether calibration in one workout correlated with calibration in the other workout. Of the fifty participants who completed both workouts, twenty-six completed the AMRAP workout first, and twenty-four completed the Rounds workout first.

I calculated bias by subtracting performance from predictions (i.e. *Predicted performance – Actual performance*). Absolute accuracy was equal to the absolute value of bias. I then calculated bias and absolute accuracy percentage scores relative to performance to account for the effects of performance value variation on calibration. The calculations were the following: (*bias/performance*) × 100 and (*absolute accuracy/performance*) × 100.

3.2.2.5 Procedure

I emailed participants with a PAR-Q health eligibility form (see Appendix 8.5.1) to check their eligibility to participate by ensuring that no previous health conditions would be exacerbated during the study. I assigned eligible participants a study slot. During each study session, participants had to first provide their informed consent. They then completed a brief experience questionnaire that asked them about their exercise background (see Appendix 8.6). After this, participants went to the gym area and warmed up for about 10 minutes. The warm-up session included all the exercises that composed the experimental workouts to ensure that participants were aware of the exercise standards before making their predictions. During the warm-up, I asked participants to indicate whether they were familiar with each exercise before taking part in the study or not (I assigned a value of one for familiar exercises, and a value of zero for unfamiliar exercises). After receiving information about the two workouts, participants gave predictions for both HIFME workouts based on how they thought, rather than how they hoped, they were going to perform. For the AMRAP workout, participants predicted the number of repetitions they would complete in ten minutes. For the Rounds workout, they predicted their finish time. Participants rested for five minutes between the two workouts. I collected ratings of perceived exertion right before and after each workout. After completing the two workouts, I debriefed participants, and the session concluded.

3.2.3 Results

3.2.3.1 Data checks

Order of workouts. Workout order did not have a significant effect on predictions, performance, bias, and absolute accuracy (all *ps* > .05; Table 3.1). In the AMRAP workout, participants made similar predictions, performed similarly, and exhibited similar bias and absolute accuracy regardless of workout order. In the Rounds workout, participants made similar predictions, and exhibited similar bias and absolute accuracy regardless of workout order. Interestingly, there was a non-significant tendency for participants to perform better when they completed the Rounds workout first, rather than second. Given that this performance pattern was not present in the AMRAP workout, it is possible that the AMRAP workout induced higher fatigue than the Rounds workout did.

Table 3.1

Data checks on AMRAP and Rounds performance, predictions, bias, and absolute accuracy when each workout was performed first or second in order.

	Completed 1st		Completed 2nd		Statistics	
Outcome Variable	***Mean***	***SD***	***Mean***	***SD***	***t***	***p***
AMRAP Performance	197 reps	41 reps	199 reps	50 reps	-0.16	.874
AMRAP Predictions	192 reps	52 reps	195 reps	60 reps	-0.18	.857
Rounds Performance	475 s	108 s	540 s	131 s	1.89	.065
Rounds Predictions	528 s	164 s	546 s	144 s	-0.43	.668
AMRAP Bias	-1.83%	21.42%	-0.53%	25.27%	-0.20	.845
AMRAP Absolute Accuracy	16.48%	13.41%	19.36%	15.74%	-0.70	.487
Rounds Bias	10.68%	23.38%	2.51%	21.50%	1.29	.205
Rounds Absolute Accuracy	20.26%	15.40%	15.88%	14.38%	1.04	.303

Note. The table compares the means and standard deviations (SD) for performance, predictions, and calibration when each workout was completed first and second.

Perceived exertion. I conducted a within subjects ANCOVA on post-workout perceived exertion to examine whether participants experienced similar exertion following each workout. I included pre-workout exertion (i.e. exertion ratings I collected right before participants started the workout) as a covariate in the analysis to control for baseline fatigue. Contrary to my assumption of equal workout difficulty, participants reported higher post-workout exertion for the AMRAP workout (M = 17.36, SE = 1.51) than the Rounds workout (M = 15.48, SE = 2.49), $F_{(1,47)} = 10.31$, $p = .002$, $\eta p^2 = .180$. Therefore, participants appeared to exert themselves more in the AMRAP than the Rounds workout.

Exercise familiarity. To examine differences in exercise familiarity between the HIFME and the non-HIFME groups prior to participation in the study, I conducted a Mann-Whitney U test. I used a non-parametric test, as familiarity scores were not normally distributed. The test illustrated that participants in the HIFME group were familiar with more of the exercises they were asked to complete (M = 5.86, SD = 0.44) than participants in the non-HIFME group (M = 5.10, SD = 1.09), U = 172.50, p = .001. This difference was consistent across workouts, as HIFME participants (AMRAP: M = 2.90, SD = 0.31; Rounds: M = 2.97, SD = 0.19) were more familiar with exercises in both the AMRAP, U = 188.00, p = .003, and the Rounds, U = 228.00, p = .012, workouts than non-HIFME participants (AMRAP: M = 2.38, SD = 0.81; Rounds: M = 2.71, SD = 0.46). Differences in familiarity with workout movements between the two groups confirmed that participants in the HIFME group had higher experience with the functional exercises used in the study than the non-HIFME group.

3.2.3.2 Performance & Predictions

To examine how experience and gender influenced performance and predictions, I conducted 2x2 between subjects ANOVAs with HIFME groups and gender as the two factors. For the relationships of months of HIFME experience and number of HIFME sessions with performance and predictions, I first conducted correlations and then multiple linear regressions where I entered both experience factors at the same time. I only conducted this analysis for data from the HIFME group. These methods of analysis were the same across workouts.

Table 3.2

Descriptive statistics for HIFME experience.

	HIFME Experience		**Non-HIFME Experience**	
Outcome Variable	***Mean***	***SD***	***Mean***	***SD***
AMRAP Performance	209 reps	50 reps	177 reps	33 reps
AMRAP Predictions	207 reps	52 reps	163 reps	59 reps
Rounds Performance	480 s	129 s	552 s	105 s
Rounds Predictions	515 s	139 s	558 s	161 s
AMRAP Bias	0.31%	16.39%	-7.48%	30.68%
AMRAP Absolute Accuracy	14.01%	8.09%	24.88%	18.81%
Rounds Bias	9.07%	22.61%	1.02%	22.53%
Rounds Absolute Accuracy	19.07%	14.84%	16.55%	14.96%

Note. The table provides information on the means and standard deviations (SD) of participants with and without HIFME experience for performance, predictions, and calibration.

3.2.3.2.1 AMRAP workout

Performance. There was a significant main effect of HIFME experience group on AMRAP performance, $F_{(1,49)} = 8.16$, $p = .006$, $\eta p^2 = .143$. Participants with HIFME experience completed more repetitions than participants without HIFME experience (Table 3.2). Surprisingly, the main effect of gender on performance was not significant, $F_{(1,49)} = 2.38$, $p = .129$, $\eta p^2 = .046$. Though male participants exhibited a tendency to complete more repetitions than female participants, this tendency did not reach significance (Table 3.3). There was no significant interaction between HIFME experience and gender on performance, $F_{(1,49)} = 1.90$, $p = .174$, $\eta p^2 = .037$.

Results from correlational analyses and the multiple regression model for AMRAP performance can be seen in Tables 3.4 and 3.5. The regression model was significant in predicting AMRAP performance, $R^2 = .464$, $F_{(2, 25)} = 10.82$, $p < .001$. Number of HIFME sessions per week was a significant predictor of performance, with participants who engaged in HIFME more frequently completing more repetitions than participants who did not participate in HIFME as frequently. Months of HIFME experience showed a similar non-significant tendency for participants who had been doing HIFME for longer to outperform participants who had been doing it for a shorter period. These results were generally consistent with correlational analyses. Overall, having a HIFME background and higher HIFME experience predicted better AMRAP performance than not having a HIFME background or having lower experience. Interestingly, gender did not have a significant influence on performance, with men exhibiting only a slight tendency to complete more repetitions than women.

Table 3.3.

Descriptive statistics for Gender.

	Male		**Female**	
Outcome Variable	***Mean***	***SD***	***Mean***	***SD***
AMRAP Performance	205 reps	49 reps	185 reps	41 reps
AMRAP Predictions	204 reps	62 reps	173 reps	53 reps
Rounds Performance	482 s	118 s	547 s	122 s
Rounds Predictions	496 s	155 s	577 s	134 s
AMRAP Bias	0.50%	26.59%	-6.54%	21.39%
AMRAP Absolute Accuracy	18.76%	18.45%	19.08%	11.14%
Rounds Bias	4.07%	25.89%	6.69%	19.17%
Rounds Absolute Accuracy	20.61%	15.73%	14.98%	13.45%

Note. The table provides information on the means and standard deviations (SD) of male and female participants for performance, predictions, and calibration.

Predictions. There was a significant main effect of HIFME experience group on AMRAP predictions, $F_{(1,49)} = 8.38$, $p = .006$, $\eta p^2 = .146$. Participants with HIFME experience predicted they would complete more repetitions than participants without HIFME experience (Table 3.2). The main effect of gender on predictions was not significant, $F_{(1,49)} = 3.95$, $p = .052$, $\eta p^2 = .075$. Male participants predicted they would complete more repetitions than female participants, but this tendency did not reach significance (Table 3.3). There was no significant interaction between HIFME experience and gender on performance, $F_{(1,49)} = 0.02$, $p = .899$, $\eta p^2 < .001$.

Table 3.4

Correlation coefficients for the associations of experience factors with the outcome variables performance and predictions.

Outcome variable	Coefficient	*r*	*p*
AMRAP Performance	Months	0.18	.346
	Sessions	0.62	< .001
AMRAP Predictions	Months	0.12	.532
	Sessions	0.41	.031
Rounds Performance	Months	-0.16	.407
	Sessions	-0.54	.002
Rounds Predictions	Months	-0.10	.593
	Sessions	-0.44	.017

Note. *r* represents the correlation coefficient of each factor with the outcome variables. *p* represents the *p*-value associated with corresponding predictor and outcome variable. Months is months of HIFME experience, and Sessions is number of HIFME sessions completed per week.

Results from correlational analyses and the multiple regression model for AMRAP predictions can be seen in Tables 3.4 and 3.5. The regression model did not predict AMRAP predictions significantly, $R^2 = .202$, $F_{(2, 25)} = 3.16$, $p = .060$. Months of HIFME experience did not contribute to performance predictions significantly. Conversely, number of HIFME sessions completed per week was a significant predictor of AMRAP predictions. Participants who engaged in HIFME training more frequently predicted they would complete more repetitions than participants who engaged in HIFME less frequently. Regression results were in line with correlational analyses. Overall, participants with a HIFME background and higher HIFME experience in terms of HIFME training sessions completed and predicted they would complete more repetitions than participants without a HIFME background or with low HIFME experience. As with performance results, gender did not have a significant influence on AMRAP predictions, with men exhibiting only a non-significant tendency to complete and predict more repetitions than women.

Table 3.5

Multiple linear regression coefficients for experience predictors on the outcome variables performance and predictions.

Outcome variable	Coefficient	*B*	*Beta*	*Std. Error*	*t*	*p*
AMRAP Performance	Months	0.46	0.28	0.24	1.91	.068
	Sessions	15.36	0.66	3.43	4.49	< .001
AMRAP Predictions	Months	0.32	0.19	0.30	1.05	.303
	Sessions	10.36	0.44	4.29	2.42	.023
Rounds Performance	Months	-1.08	-0.26	0.67	-1.61	.119
	Sessions	-35.22	-0.59	9.59	-3.68	.001
Rounds Predictions	Months	-0.82	-0.18	0.80	-1.03	.312
	Sessions	-30.46	-0.47	11.38	-2.68	.013

Note. Factors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. Months is months of HIFME experience, and Sessions is number of HIFME sessions completed per week.

3.2.3.2.2 Rounds workout

Performance. There was a significant main effect of HIFME experience on Rounds performance, $F_{(1,52)} = 6.51$, $p = .014$, $\eta p^2 = .111$. Participants with HIFME experience finished the workout faster than participants without HIFME experience (Table 3.2). There was also a significant main effect of gender, $F_{(1,52)} = 5.43$, $p = .024$, $\eta p^2 = .095$, such that male participants finished the Rounds workout faster than female participants (Table 3.3). There was no significant interaction between HIFME experience and gender on performance, $F_{(1,52)} = 0.78$, $p = .383$, $\eta p^2 = .015$.

Results from correlational analyses and the multiple regression model for Rounds performance can be seen in Tables 3.4 and 3.5. The regression model significantly predicted Rounds performance, $R^2 = .358$, $F_{(2, 26)} = 7.26$, $p = .003$. Consistent with correlational analyses, participants who engaged in HIFME more frequently during the week were faster to complete the workout than participants who did not participate in HIFME as frequently. Though participants who had been engaging in HIFME for longer tended to outperform those who had been engaging in HIFME for a shorter period, this tendency was not significant.

Overall, experienced and male participants outperformed less experienced and female participants respectively.

Predictions. There was no significant main effect of HIFME experience on Rounds predictions, $F_{(1,52)} = 1.76$, $p = .191$, $\eta p^2 = .033$. Participants with HIFME experience made similar performance predictions as participants without HIFME experience (Table 3.2). Conversely, the main effect of gender on Rounds predictions was significant, $F_{(1,52)} = 4.71$, $p = .035$, $\eta p^2 = .083$. Male participants predicted they would finish the Rounds workout faster than female participants (Table 3.3). There was no significant interaction between HIFME experience and gender on performance, $F_{(1,52)} = 1.10$, $p = .300$, $\eta p^2 = .021$.

Results from correlational analyses and the multiple regression model for Rounds predictions can be seen in Tables 3.4 and 3.5. The regression model significantly predicted Rounds predictions, $R^2 = .224$, $F_{(2, 26)} = 3.76$, $p = .037$. Consistent with correlational analyses, participants who engaged in HIFME more frequently during the week predicted they would be faster to complete the workout than participants who did not participate in HIFME as frequently. Months of HIFME experience were not associated with performance predictions for the Rounds workout. Overall, only gender and HIFME sessions per week in the HIFME group contributed to predictions in the Rounds workout.

3.2.3.3 Calibration

To examine the influence of each factor on calibration, I conducted correlational analyses, and multiple linear regressions, where I entered all predictors at the same time to account for shared variance. In the first regression model, the factors I entered were HIFME experience and gender. In the second model, I also added performance as a measure of expertise. Similar to Section 2.2.3.4, I did this because performance was associated with experience and gender (see Section 3.2.3.2), and I wanted to investigate the influence of each factor on calibration with and without the inclusion of performance. I followed the same methodology in analyses for the HIFME group alone when I examined experience factors that were exclusive to participants with a HIFME background. I repeated these analyses for bias and absolute accuracy across workouts.

3.2.3.3.1 AMRAP workout

Bias. In the AMRAP workout, positive bias values indicate overconfidence, i.e. performing worse than expected, whilst negative values indicate underconfidence, i.e. performing better than expected. Values close to zero indicate lack of bias. Generally, participants made unbiased predictions for the AMRAP workout, exhibiting only a minor and non-significant tendency towards underconfidence (*M* = -3.22%, *SD* = 24.00%), $t_{(52)} = -0.98$, $p = .333$, $d = -0.13$.

Table 3.6

Correlation coefficients for the associations of demographic factors with the outcome variables AMRAP bias and absolute accuracy.

		Bias		Abs Acc	
Analysed sample	**Factor**	***r***	***p***	***r***	***p***
Both groups	HIFME_exp	-0.16	.243	0.37	.007
	Gender	0.15	.291	-0.01	.939
	Performance	-0.09	.510	-0.20	.153
HIFME only	Months	-0.13	.508	-0.12	.550
	Sessions	-0.28	.151	0.21	.277
	Performance	-0.34	.068	-0.04	.826

Note. *r* represents the correlation coefficient of each factor with the outcome variables. *p* represents the *p*-value associated with corresponding predictor and outcome variable. HIFME_exp refers to HIFME group membership (HIFME_exp = 0, if the participant had a HIFME background; HIFME_exp = 1, if the participant did not have a HIFME background), Gender to participant gender (Gender = 0, if female; Gender = 1, if male), Months to months of HIFME experience, and Sessions to number of HIFME sessions completed per week.

Results from correlational analyses and the multiple regression models for AMRAP bias can be seen in Tables 3.6 and 3.7. Neither regression model was significant in predicting AMRAP bias across experience groups, regardless of whether they included performance, $R^2 = .087$, $F_{(3, 49)} = 1.57$, $p = .210$, or not, $R^2 = .047$, $F_{(2, 50)} = 1.24$, $p = .298$. In the same vein, no factor was significant in predicting bias. In the model including performance, participants in the non-HIFME group showed a tendency to be underconfident, whereas the HIFME group was unbiased (Fig. 3.6A; Table 3.2); women were more likely than men to be underconfident (Fig 3.6C; Table 3.3); and better performance was associated with less overconfidence/more underconfidence (Fig. 3.6E). However, none of these relationships were significant, suggesting that there is not sufficient evidence to support them. In the regression analyses examining experience in the HIFME group alone, months of HIFME experience and HIFME training frequency also failed to significantly predict AMRAP bias regardless of whether performance was included in the model, $R^2 = .134$, $F_{(3, 24)} = 1.24$, $p = .317$, or not, $R^2 = .108$,

$F_{(2, 25)} = 1.52$, $p = .239$ (Fig. 3.7A, 3.7C, & 3.7E). Overall, the above analyses suggest that expertise, experience, and gender were not significant predictors of AMRAP bias—I observed similar results in the correlational analyses where I examined the individual association of each factor with bias.

Table 3.7

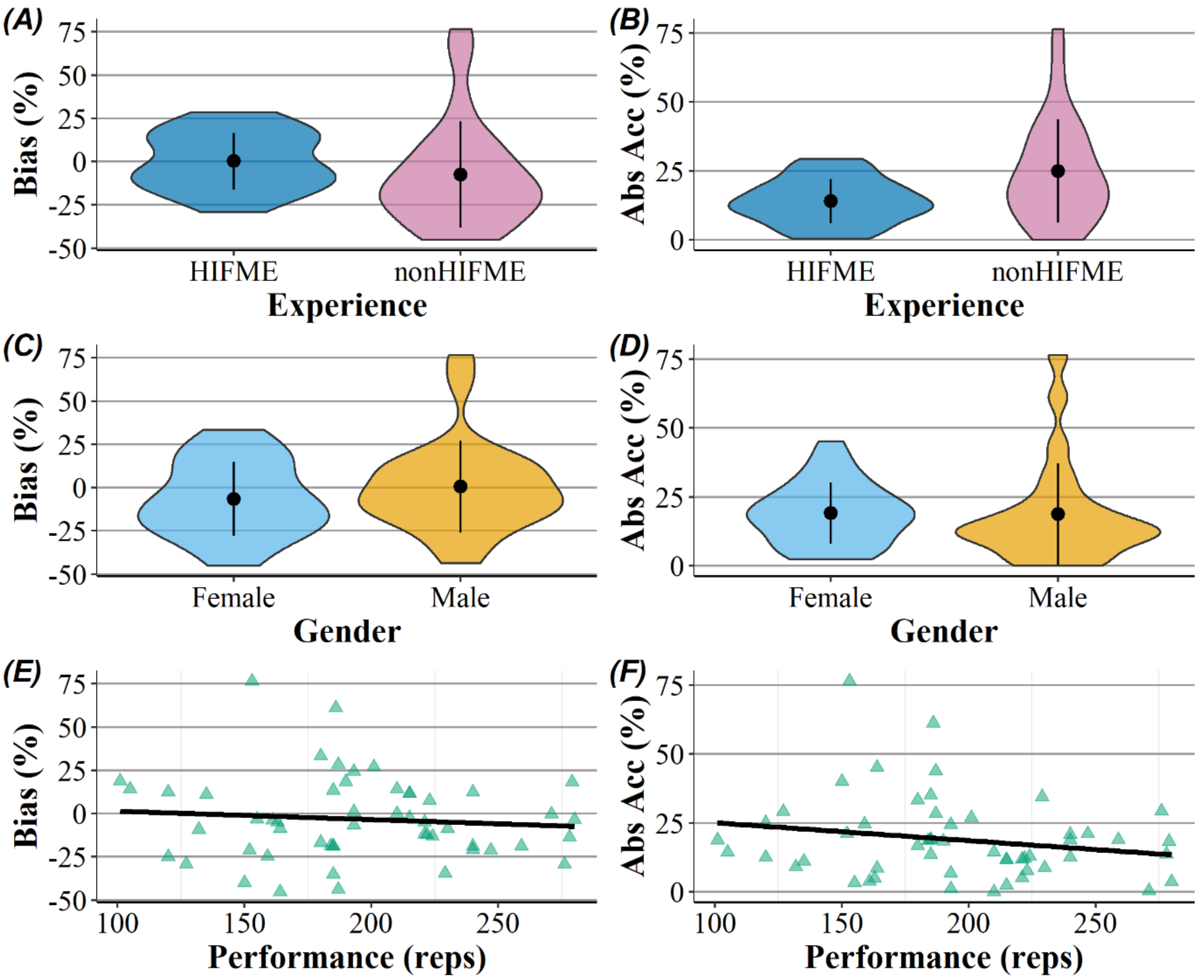
Multiple regression coefficients for demographic predictors on the outcome variable AMRAP bias.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No Performance	Intercept	4.62	—	10.66	0.43	.667
	HIFME_exp	-7.62	-0.16	6.60	-1.16	.253
	Gender	6.86	0.14	6.58	1.04	.302
Performance included	Intercept	31.55	—	21.16	1.49	.142
	HIFME_exp	-11.35	-0.24	7.00	-1.62	.111
	Gender	9.11	0.19	6.68	1.36	.179
	Performance	-0.12	-0.22	0.08	-1.47	.148
No Performance – HIFME only	Intercept	13.16	—	8.42	1.56	.131
	Months	-0.09	-0.18	0.10	-0.92	.364
	Sessions	-2.23	-0.31	1.39	-1.60	.122
Performance included – HIFME only	Intercept	21.57	—	13.02	1.66	.111
	Months	-0.06	-0.11	0.11	-0.56	.583
	Sessions	-1.16	-0.16	1.88	-0.62	.544
	Performance	-0.07	-0.22	0.08	-0.85	.404

Note. Factors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. HIFME_exp refers to HIFME group membership (HIFME_exp = 0, if the participant had a HIFME background; HIFME_exp = 1, if the participant did not have a HIFME background), Gender to participant gender (Gender = 0, if female; Gender = 1, if male), Months to months of HIFME experience, and Sessions to number of HIFME sessions completed per week.

Figure 3.6

Plots illustrating the associations between demographic factors and AMRAP bias and absolute accuracy.



Note. Panels A and C show violin plots that illustrate the influence of HIFME experience and gender on AMRAM bias. Panels B and D show violin plots that illustrate the influence of HIFME experience and gender on AMRAP absolute accuracy. Panels E and F show scatterplots that illustrate the influence of performance on AMRAP bias and absolute accuracy respectively. The perimeter of each violin plot illustrates density, the central point represents the mean, and the vertical line represents +/- one standard deviation. Reps stands for repetitions completed in the AMRAP workout.

Absolute Accuracy. Values closer to zero indicate high precision, whereas larger values indicate poor precision. Results from correlational analyses and the multiple regression models for AMRAP absolute accuracy can be seen in Tables 3.6 and 3.8. The regression model that did not include finish time was significant in predicting absolute accuracy in the analyses using data from both experience groups, $R^2 = .135$, $F_{(2, 50)} = 3.89$, $p = .027$. However, the model that included performance was not significant, $R^2 = .140$, $F_{(3, 49)} = 2.66$, $p = .058$.

Across models, participants with HIFME experience made more precise predictions for the AMRAP workout than participants without HIFME experience (Fig. 3.6B; Table 3.2). Gender and performance did not contribute to absolute accuracy (Fig. 3.6D & 3.6F; Table 3.3). When examining absolute accuracy in the HIFME group alone, neither regression model was significant in predicting absolute accuracy regardless of whether performance was included, $R^2 = .087$, $F_{(3, 24)} = 0.76$, $p = .525$, or not, $R^2 = .052$, $F_{(2, 25)} = 0.69$, $p = .510$. Months of HIFME experience (Fig. 3.7B), training frequency (Fig. 3.7D), and performance (Fig. 3.7F) all failed to predict absolute accuracy. Overall, the only significant result of the analysis was that participants with HIFME experience were more precise in their predictions than participants without HIFME experience. As with bias analyses, regression findings for absolute accuracy were consistent with correlational analyses of individual associations (Table 3.5).

Table 3.8

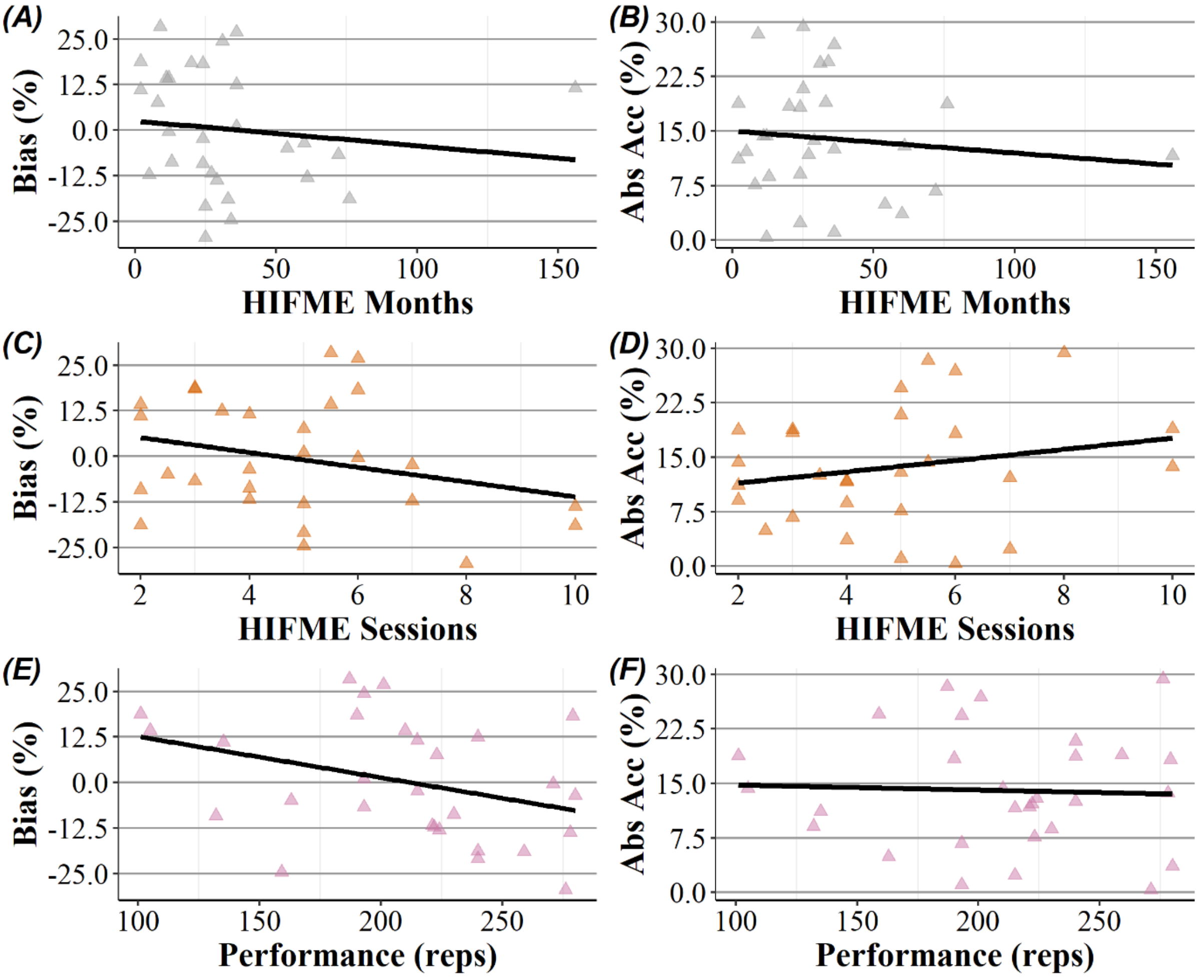
Multiple regression coefficients for demographic predictors on the outcome variable AMRAP absolute accuracy.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No Performance	Intercept	3.16	—	6.30	0.50	.618
	HIFME_exp	10.87	0.37	3.90	2.79	.007
	Gender	-0.05	0.00	3.89	-0.01	.989
Performance included	Intercept	9.32	—	12.74	0.73	.468
	HIFME_exp	10.02	0.34	4.21	2.38	.021
	Gender	0.46	0.02	4.02	0.12	.909
	Performance	-0.03	-0.08	0.05	-0.56	.580
No Performance – HIFME only	Intercept	10.86	—	4.33	2.51	.019
	Months	-0.02	-0.09	0.05	-0.44	.667
	Sessions	0.73	0.20	0.72	1.01	.321
Performance included – HIFME only	Intercept	15.70	—	6.67	2.35	.027
	Months	0.00	-0.01	0.05	-0.07	.948
	Sessions	1.34	0.37	0.96	1.39	.177
	Performance	-0.04	-0.26	0.04	-0.96	.349

Note. Factors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. HIFME_exp refers to HIFME group membership (HIFME_exp = 0, if the participant had a HIFME background; HIFME_exp = 1, if the participant did not have a HIFME background), Gender to participant gender (Gender = 0, if female; Gender = 1, if male), Months to months of HIFME experience, and Sessions to number of HIFME sessions completed per week.

Figure 3.7

Scatter plots illustrating the associations between experience factors and AMRAP bias and absolute accuracy for the HIFME group.



Note. Panels A, C, and E illustrate the influence of HIFME months of experience, sessions per week, and performance on AMRAP bias for the HIFME group alone. Panels B, D, and F illustrate the influence of the same factors on AMRAP absolute accuracy. Reps stands for repetitions completed in the AMRAP workout.

3.2.3.3.2 Rounds workout

Bias. In the Rounds workout, positive bias values indicate underconfidence, whilst negative values indicate overconfidence. Values close to zero indicate a lack of bias. Generally, participants across experience groups exhibited a non-significant tendency to be underconfident in their predictions for the Rounds workout (*M* = 5.33%, *SD* = 22.73%), $t_{(55)}$ = 1.76, *p* = .085, *d* = 0.24.

Table 3.9

Correlation coefficients for the associations of demographic factors with the outcome variables Rounds bias and absolute accuracy.

		Bias		**Abs Acc**	
Analysed sample	**Factor**	***r***	***p***	***r***	***p***
Both groups	HIFME_exp	-0.18	.189	-0.09	.531
	Gender	-0.06	.670	0.19	.157
	Performance	-0.21	.122	-0.19	.157
HIFME only	Months	0.04	.835	-0.30	.112
	Sessions	0.11	.564	0.23	.240
	Performance	-0.32	.087	-0.08	.667

Note. *r* represents the correlation coefficient of each factor with the outcome variables. *p* represents the *p*-value associated with corresponding predictor and outcome variable. HIFME_exp refers to HIFME group membership (HIFME_exp = 0, if the participant had a HIFME background; HIFME_exp = 1, if the participant did not have a HIFME background), Gender to participant gender (Gender = 0, if female; Gender = 1, if male), Months to months of HIFME experience, and Sessions to number of HIFME sessions completed per week.

Results from correlational analyses and the multiple regression models for Rounds bias can be seen in Tables 3.9 and 3.10. When examining the overall sample, neither regression model was significant in predicting bias, regardless of whether they included performance, $R^2 = .068$, $F_{(3, 52)} = 1.26$, $p = .298$, or not, $R^2 = .033$, $F_{(2, 53)} = 0.91$, $p = .408$. Accordingly, no factor was a significant contributor of Rounds bias (see Fig. 3.8A, 3.8C, & 3.8E; Tables 3.2 & 3.3). Slower performers only showed a non-significant tendency to be more overconfident/less underconfident than faster runners. Similarly, the regression models on Rounds bias for the HIFME group alone were not significant with, $R^2 = .102$, $F_{(3, 25)} = 0.95$, $p = .432$, and without, $R^2 = .016$, $F_{(2, 26)} = 0.21$, $p = .809$, the inclusion of performance (Fig. 3.9A, 3.9C, & 3.9E). No factor was a significant predictor of bias, and, as with correlational analyses on the individual associations between demographic factors and bias, lower performance was only non-significantly associated with higher overconfidence/lower underconfidence. Overall, no factor predicted bias in the Rounds workout across analyses.

Table 3.10

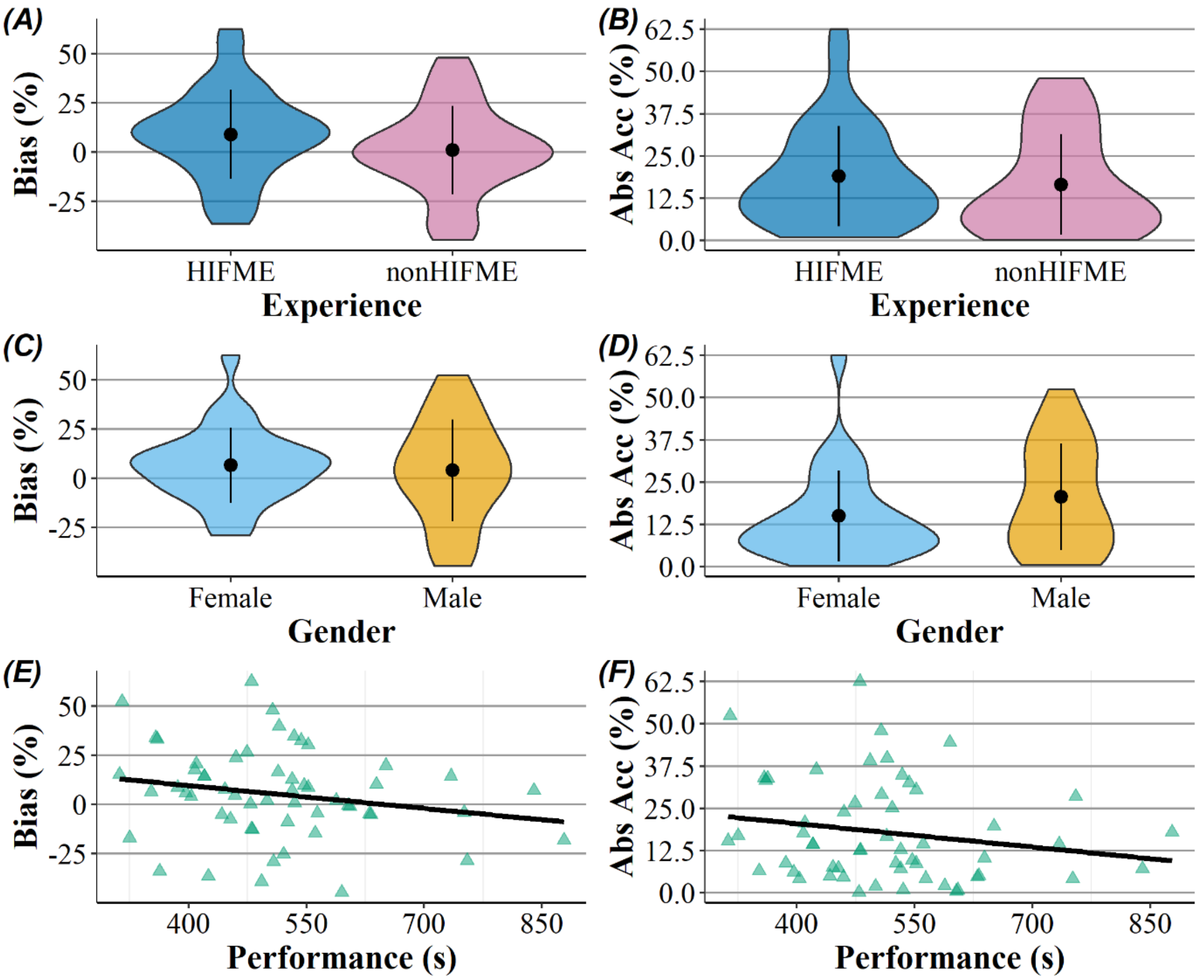
Multiple regression coefficients for demographic predictors on the outcome variable Rounds bias.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No Performance	Intercept	17.74	—	9.69	1.83	.073
	HIFME_exp	-7.85	-0.17	6.14	-1.28	.206
	Gender	-1.76	-0.04	6.13	-0.29	.775
Performance included	Intercept	34.18	—	15.25	2.24	.029
	HIFME_exp	-4.83	-0.11	6.46	-0.75	.459
	Gender	-4.56	-0.10	6.40	-0.71	.480
	Performance	-0.04	-0.21	0.03	-1.39	.171
No Performance – HIFME only	Intercept	0.95	—	12.64	0.08	.941
	Months	0.05	0.06	0.15	0.31	.759
	Sessions	1.28	0.12	2.08	0.62	.542
Performance included – HIFME only	Intercept	45.08	—	31.05	1.45	.159
	Months	-0.02	-0.03	0.15	-0.16	.872
	Sessions	-0.97	-0.09	2.49	-0.39	.700
	Performance	-0.06	-0.37	0.04	-1.55	.134

Note. Factors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. HIFME_exp refers to HIFME group membership (HIFME_exp = 0, if the participant had a HIFME background; HIFME_exp = 1, if the participant did not have a HIFME background), Gender to participant gender (Gender = 0, if female; Gender = 1, if male), Months to months of HIFME experience, and Sessions to number of HIFME sessions completed per week.

Figure 3.8

Plots illustrating the associations between demographic factors and Rounds bias and absolute accuracy.



Note. Panels A and C show violin plots that illustrate the influence of HIFME experience and gender on Rounds bias. Panels B and D show violin plots that illustrate the influence of HIFME experience and gender on Rounds absolute accuracy. Panels E and F show scatterplots that illustrate the influence of performance on Rounds bias and absolute accuracy respectively. The perimeter of each violin plot illustrates density, the central point represents the mean, and the vertical line represents +/- one standard deviation.

Absolute Accuracy. Results from correlational analyses and the multiple regressions for Rounds absolute accuracy can be seen in Tables 3.9 and 3.11. The regression models used to predict precision across experience groups were not significant with, $R^2 = .062$, $F_{(2, 52)} = 1.14$, $p = .341$, and without, $R^2 = .048$, $F_{(2, 53)} = 1.34$, $p = .270$, the inclusion of performance. HIFME experience (Fig. 3.8B; Table 3.2), gender (Fig. 3.8D; Table 3.3), and performance (Fig. 3.8F) all failed to predict absolute accuracy for the Rounds workout. When examining the HIFME group alone, the regression models also failed to predict absolute accuracy regardless of whether performance was included, $R^2 = .120$, $F_{(3, 25)} = 1.14$, $p = .353$, or not, $R^2 = .119$, $F_{(2, 26)} = 1.76$, $p = .192$, and no factor was a significant predictor of absolute accuracy (Fig. 3.9B, 3.9D, & 3.9F). Athletes with more months of HIFME experience were more precise than runners with less experience, but this relationship did not reach significance. Overall, no factor predicted absolute accuracy in the Rounds workout, mirroring correlational analyses.

Table 3.11

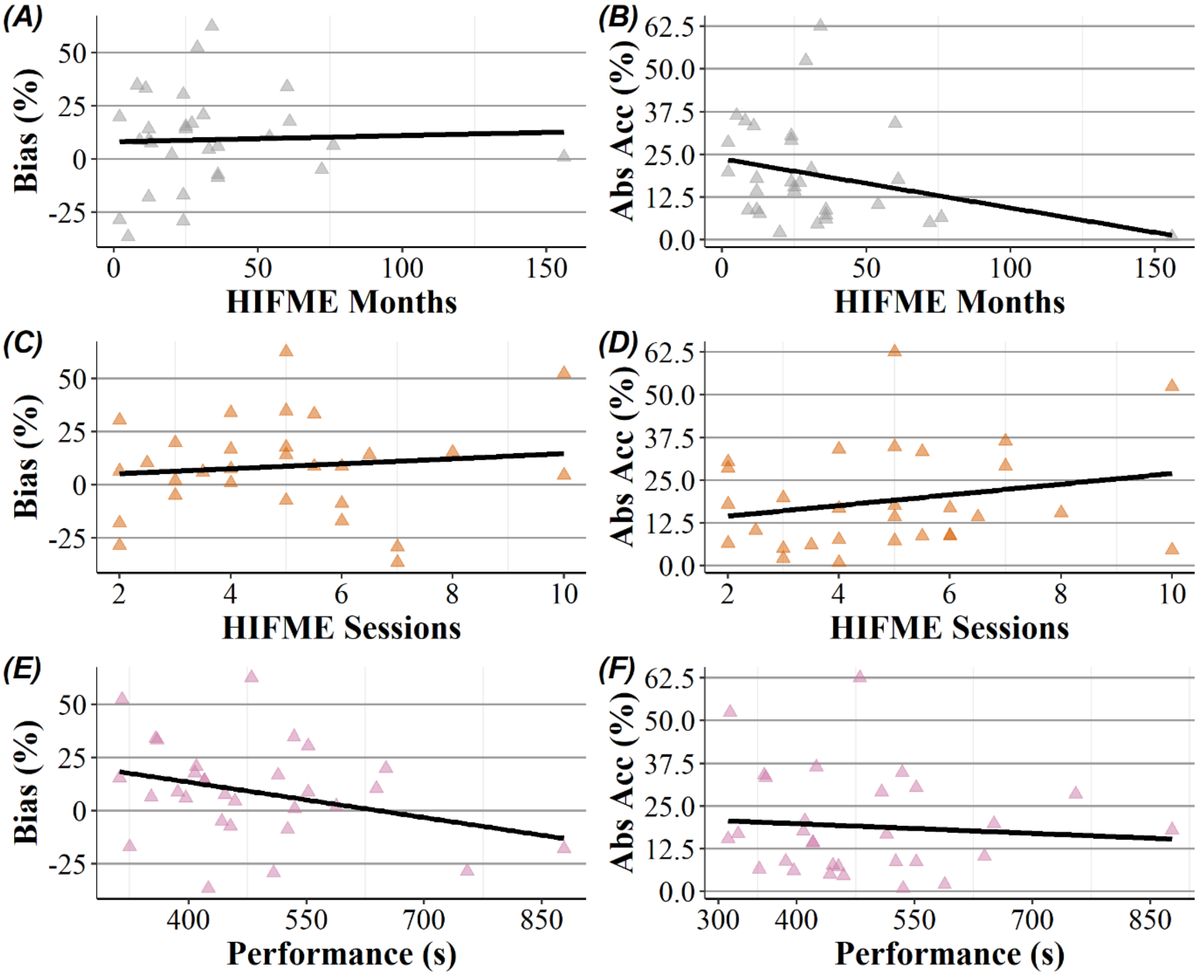
Multiple regression coefficients for demographic predictors on the outcome variable Rounds absolute accuracy.

	Coefficient	***B***	***Beta***	***Std. Error***	***t***	***p***
No Performance	Intercept	19.45	—	6.27	3.11	.003
	HIFME_exp	-3.18	-0.11	3.97	-0.80	.427
	Gender	5.98	0.20	3.96	1.51	.137
Performance included	Intercept	26.16	—	9.97	2.62	.011
	HIFME_exp	-1.94	-0.07	4.23	-0.46	.648
	Gender	4.84	0.17	4.18	1.16	.253
	Performance	-0.02	-0.13	0.02	-0.87	.391
No Performance – HIFME only	Intercept	17.07	—	7.88	2.17	.040
	Months	-0.13	-0.27	0.09	-1.42	.167
	Sessions	1.25	0.18	1.30	0.97	.342
Performance included – HIFME only	Intercept	20.09	—	20.27	0.99	.331
	Months	-0.13	-0.28	0.10	-1.38	.180
	Sessions	1.10	0.16	1.63	0.68	.506
	Performance	0.00	-0.04	0.03	-0.16	.872

Note. Factors were entered at the same time in each multiple regression model. *B* and *Beta* represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. HIFME_exp refers to HIFME group membership (HIFME_exp = 0, if the participant had a HIFME background; HIFME_exp = 1, if the participant did not have a HIFME background), Gender to participant gender (Gender = 0, if female; Gender = 1, if male), Months to months of HIFME experience, and Sessions to number of HIFME sessions completed per week.

Figure 3.9

Scatter plots illustrating the associations between experience factors and Rounds bias and absolute accuracy for the HIFME group.



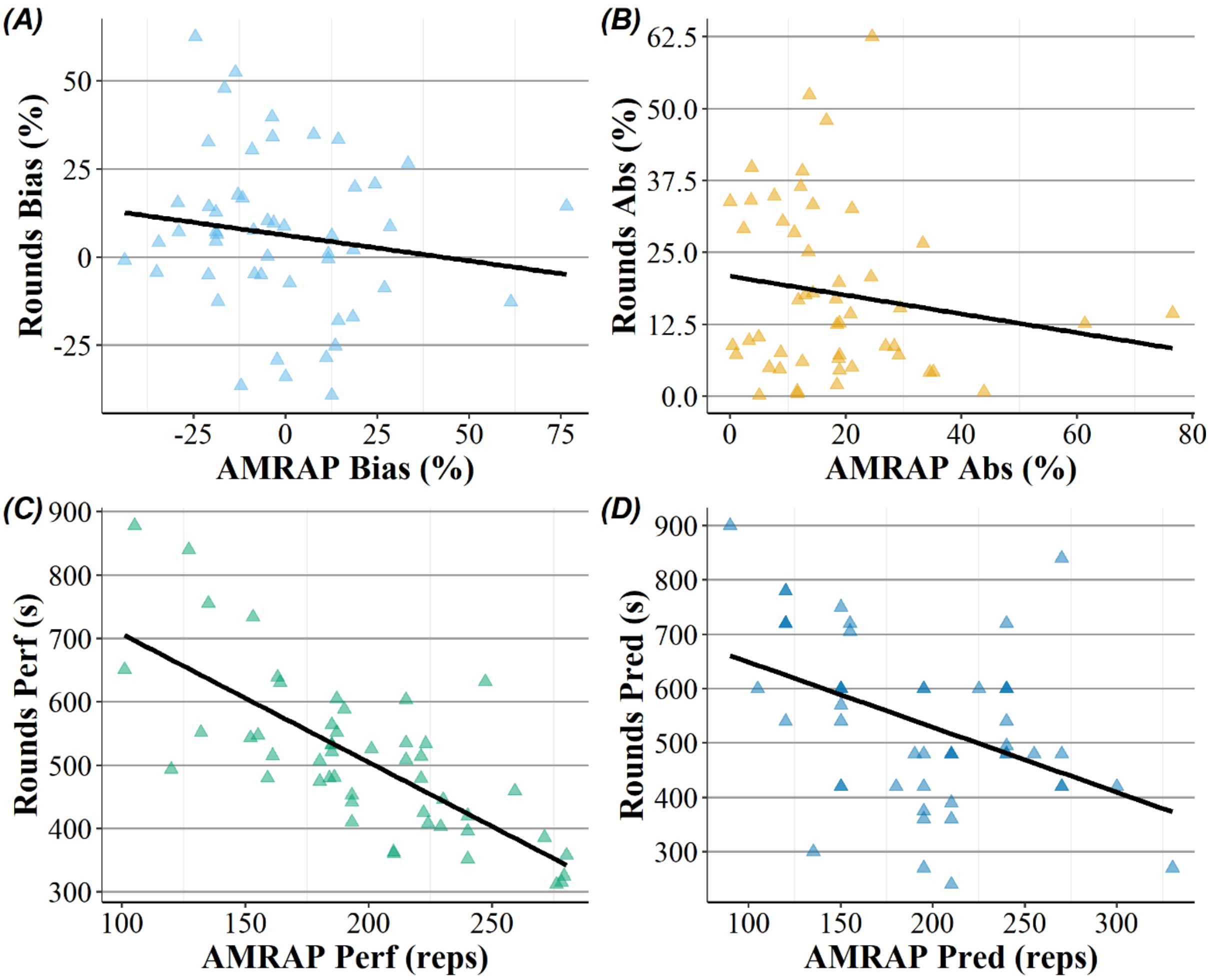
Note. Panels A, C, and E illustrate the influence of HIFME months of experience, sessions per week, and performance on Rounds bias for the HIFME group alone. Panels B, D, and F illustrate the influence of the same factors on Rounds absolute accuracy.

3.2.3.3.3 Workout comparisons

I conducted correlations between the AMRAP and the Rounds workouts for bias and absolute accuracy to examine whether we can use calibration information from one workout to make predictions about another. AMRAP bias scores did not correlate with Rounds bias scores (Fig. 3.10A), $r_{(48)} = -.15$, $p = .302$. Similarly, AMRAP absolute accuracy scores did not correlate with Rounds absolute accuracy (Fig. 3.10B), $r_{(48)} = -.16$, $p = .269$. Therefore, the present findings suggest that it might be difficult or impossible to use calibration from one HIFME workout to make inferences about calibration in another HIFME workout.

Figure 3.10

Scatterplots illustrating correlations between the two HIFME workouts for calibration measures, performance, and predictions.

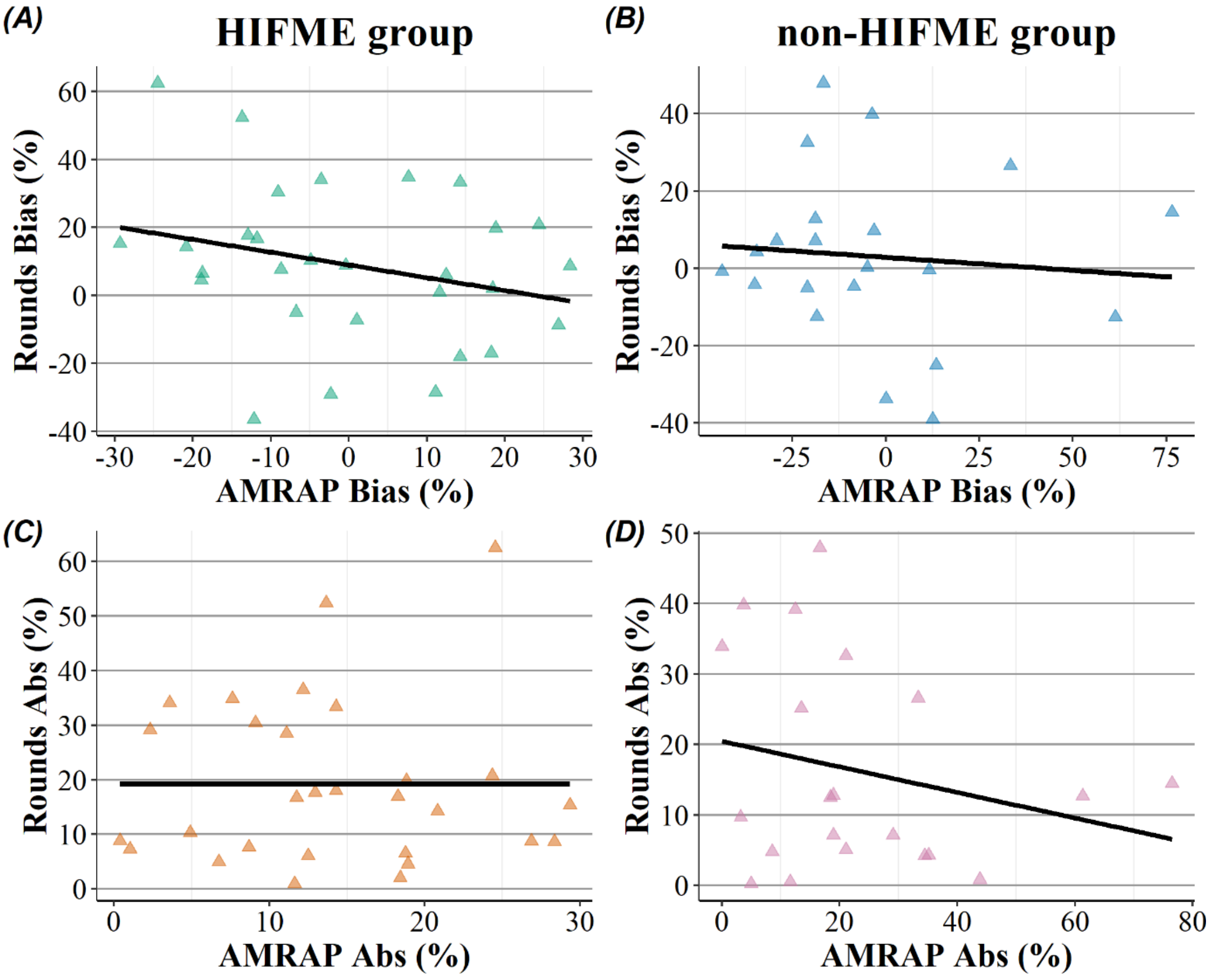


Note. Panel A shows the correlation between AMRAP and Rounds bias. Panel B shows the correlation between AMRAP and Rounds absolute accuracy. Panel C shows the correlation between AMRAP and Rounds performance. Panel D shows the correlation between AMRAP and Rounds predictions. Reps stands for the number of repetitions completed in the AMRAP workout.

Conversely, there was a strong negative correlation for performance between the two workouts (Fig. 3.10C), $r_{(48)} = -.74$, $p < .001$, indicating that good performers in the AMRAP workout tended to also be good performers in the Rounds workout. The correlation between workout predictions was weaker, but in the same direction (Fig. 3.10D), $r_{(48)} = -.43$, $p = .002$, suggesting that individuals who predicted more repetitions for the AMRAP workout were more likely to also predict faster finish times for the Rounds workout. However, the relationship between predictions from the two workouts was not strong; thereby, factors other than performance capacity seem to influence how participants estimate prospective performance.

Figure 3.11

Scatterplots illustrating correlations between the two HIFME workouts within each experience group for bias and absolute accuracy.

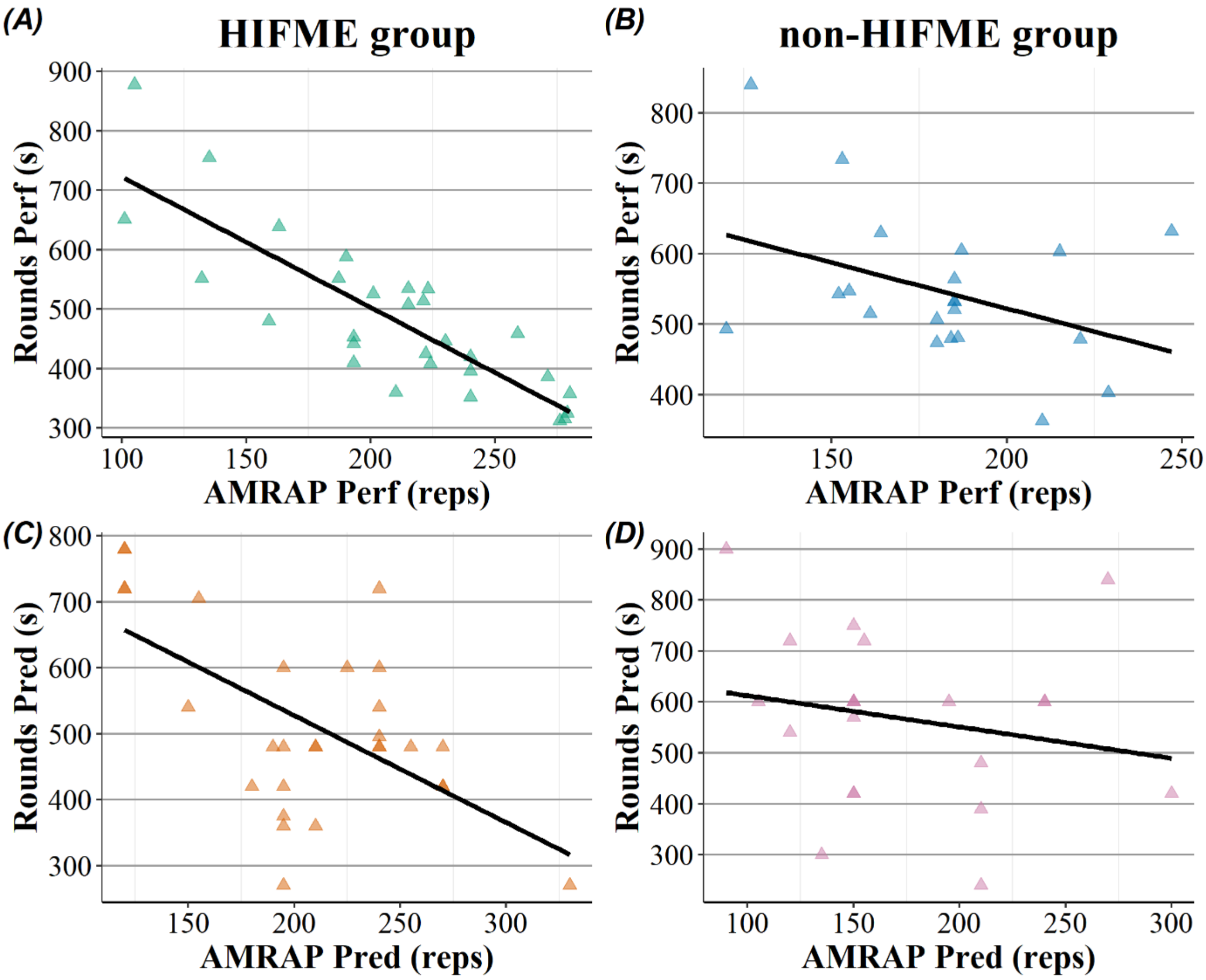


Note. Panel A shows the correlation between AMRAP and Rounds bias for the HIFME group. Panel B shows the correlation between AMRAP and Rounds bias for the non-HIFME group. Panel C shows the correlation between AMRAP and Rounds absolute accuracy for the HIFME group. Panel D shows the correlation between AMRAP and Rounds absolute accuracy for the non-HIFME group.

I was also interested in the extent to which HIFME experience influenced the previous associations. Within the HIFME group, there was a small and non-significant negative correlation for bias between the two workouts, $r_{(27)} = -.27$, $p = .161$. Overconfident and underconfident athletes in the AMRAP workout were also slightly more likely to be overconfident and underconfident respectively in the Rounds workout (Fig. 3.11A). There was no correlation for bias between the two workouts in the non-HIFME group (Fig. 3.11B), $r_{(19)} = -.09$, $p = .690$. There was also no correlation on absolute accuracy between the two workouts in either the HIFME group (Fig. 3.11C), $r_{(27)} = .00$, $p = 1.000$, or the non-HIFME group (Fig. 3.11D), $r_{(19)} = -.24$, $p = .302$. Correlations on performance and predictions between the two workouts were significant and moderate-to-strong for the HIFME group (Fig. 3.12A & 3.12C), performance: $r_{(27)} = -.84$, $p < .001$; predictions: $r_{(27)} = -.59$, $p = .001$, compared to the non-HIFME group, where they were non-significant and weak (Fig. 3.12B & 3.12D), performance: $r_{(19)} = -.39$, $p = .079$; predictions: $r_{(19)} = -.20$, $p = .374$. Though these results are preliminary, they illustrate that whether athletes perform and make predictions consistently across different workouts may depend on factors such as experience.

Figure 3.12

Scatterplots illustrating correlations between the two HIFME workouts within each experience group for performance and predictions.



Note. Panel A shows the correlation between AMRAP and Rounds performance for the HIFME group. Panel B shows the correlation between AMRAP and Rounds performance for the non-HIFME group. Panel C shows the correlation between AMRAP and Rounds predictions for the HIFME group. Panel D shows the correlation between AMRAP and Rounds predictions for the non-HIFME group. Reps stands for the number of repetitions completed in the AMRAP workout.

3.3 DISCUSSION

In the present chapter, I examined the associations of expertise, experience, and gender with calibration in a complex and unpredictable exercise modality, using data collected from two HIFME workouts. I anticipated that athletes with high expertise and a HIFME background would be better calibrated than athletes with low expertise and without a HIFME background respectively. I also anticipated that more experienced HIFME athletes would be better calibrated than less experienced HFIME athletes. For gender, I expected female athletes to be less overconfident or more underconfident than male athletes.

Surprisingly, expertise did not contribute to HIFME calibration. The only associations observed (i.e. good performers being more underconfident and precise compared to the more overconfident and less precise poor performers across workouts) were descriptive, non-significant, and inconsistent across analyses. These results are inconsistent with findings from running studies in Chapter 2 and Krawczyk and Wilamowski (2016), as well as physical education research (Kolovelonis, 2019; Kolovelonis & Goudas, 2018), where better performance has been consistently associated with higher underconfidence/lower overconfidence and higher precision than poorer performance. They are instead in line with golf and tennis studies, where participants were categorised in experts and non-experts based on non-performance criteria (Fogarty & Else, 2005; Fogarty & Ross, 2007), and where expertise had a limited influence on calibration. It is not clear why performance did not exhibit the expected pattern of results in Study 3. A possible explanation is low power resulting from recruiting a moderate sample size, as exercise studies assessing expertise based on performance have typically recruited more participants (e.g. more than 100 athletes). Given that the non-significant patterns of expertise influence on calibration observed here were congruent with the results of these studies, it is possible that more participants were just required to observe the relationship reliably.

In accordance with my prediction, participants with HIFME experience were more precise than participants without HIFME experience, but only for the AMRAP workout. Experience did not make a significant contribution to any other calibration measure in either workout. Only two descriptive associations were observed: the HIFME and the non-HIFME groups were unbiased and underconfident respectively in the AMRAP workout; and more months of HIFME experience were associated with higher precision compared to fewer months of experience for the Rounds workout when examining the HIFME group. Nonetheless, these associations were not significant, and thus cannot be used to infer a relationship between experience and calibration. The finding that HIFME experience did not contribute to calibration in the Rounds workout in the same way it did in the AMRAP workout was surprising, but could simply result from differences in workout format and exercises used. Overall, the present findings were only partially in line with previous research illustrating a positive influence of experience on calibration (e.g. Kolovelonis, 2019). This supports my suggestion in Chapter 2 that experience associations with calibration are not always strong and can be inconsistent. Though exercise experience does appear to contribute to better calibration even in complex and unpredictable modalities, such as HIFME, the extent to which this relationship is reliable is unclear, and thus warrants further examination.

Athletes of either gender were similarly calibrated across workouts, as the only differences observed were descriptive. Though females athletes were more underconfident than male athletes in the AMRAP workout, and male athletes were less precise than female athletes in the Rounds workout, these findings were inconsistent across analyses and non-significant, and thus cannot be used as evidence for the presence of gender differences in HIFME calibration. Results from the present study are inconsistent with findings from Chapter 2 and other running calibration studies (Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018), where there was evidence of male runners being more overconfident (or less underconfident) than female runners. Instead, they are congruent with physical education studies that have observed a lack of gender differences in basketball shooting and dribbling calibration (Kolovelonis, 2019; Kolovelonis & Goudas, 2018; Kolovelonis, Goudas, & Dermitzaki, 2012). However, it is not clear whether this is a reliable finding that illustrates a lack of gender influence on HIFME calibration or simply the result of an insufficient sample size. If HIFME is perceived to be a gender-neutral sport, then gender typedness (i.e. whether a task is perceived to be masculine, gender-neutral, or feminine) could affect whether there are gender differences in HIFME bias. This is because women might show considerably less confidence than men in masculine, but not necessarily gender-neutral or feminine, tasks (e.g. Jakobsson et al., 2013). However, running is generally considered to be a gender-neutral sport (Alvariñas-Villaverde, López-Villar, Fernández-Villarino, & Alvarez-Esteban, 2017; Sobal & Milgrim, 2019; Xiang, McBride, Lin, Gao, & Francis, 2018), and male runners still tend to exhibit higher overconfidence/lower underconfidence than female runners. Therefore, the extent to which factors such as gender typedness or sample size led to the present null findings needs to be examined by conducting more, similar research.

Collecting calibration data from two different HIFME workouts allowed me to investigate whether athletes would exhibit consistent calibration across different tasks. This is of particular interest in HIFME, where athletes constantly engage in workouts that consist of different exercises and formats, and calibration patterns might thus not be as consistent across workouts as they might be in less variable exercise modalities (e.g. running). Interestingly, there were no correlations for bias and absolute accuracy between the AMRAP and the Rounds workouts. Consequently, calibration from one workout could not predict calibration in the other workout. Conversely, there was a strong correlation for performance, suggesting that high physical fitness contributed to better performance in both workouts. Similarly, athletes who predicted good performance for the AMRAP workout tended to also predict good performance for the Rounds workout, though this correlation was small-to-moderate. This suggests that physical fitness was not as strong a contributor to performance predictions as it was to performance. Other factors must have also affected how participants made their performance estimates, which could include workout format and exercise movement experience.

Analysis of the HIFME group showed stronger correlations for predictions and performance between the two workouts than the collapsed data previously examined. In contrast, the non-HIFME group showed small and non-significant correlations for performance predictions and performance respectively. Therefore, HIFME experience appears to influence whether athletes make predictions and perform consistently across different HIFME workouts. Since HIFME implements a wide range of exercises, athletes without HIFME experience might not be similarly familiar with each workout, leading to inconsistencies in predictions and performance. Such inconsistences can in turn create difficulties in setting appropriate performance goals, implementing effective training and competition strategies, and remaining motivated. Nonetheless, it should be mentioned that very rigid assessments of performance could also have a negative effect on HIFME self-regulation, with athletes failing to identify areas where they need to improve. Because of this, the moderate correlation between workouts predictions observed in the HIFME group could be optimal in ensuring effective self-regulation.

The present findings on HIFME calibration have important practical and theoretical implications. The influence of HIFME experience on AMRAP precision suggests that athletes who wish to transition to HIFME need to take into consideration that previous experience with other types of exercise does not necessitate high metacognitive awareness of

HIFME performance. Given that participants in the HIFME group had more familiarity with the workout movements, it is likely important for athletes new to HIFME to quickly become familiar with the exercises they typically perform, as it can influence their performance estimates. Coaches should thus ensure that athletes have practiced the exercises they have to perform during workouts sufficiently, as doing so can ultimately contribute to performance optimisation and injury risk reduction.

The lack of associations between other experience factors and calibration is also interesting, as it highlights potential challenges in making predictions for HIFME workouts. Even experienced athletes might not be well-calibrated enough for their difference with inexperienced athletes to show given the sample size used in the present study. HIFME athletes should thus monitor and evaluate their performance constantly, and avoid assuming that their experience is sufficient to ensure good calibration across workouts. This is further supported by the absence of significant correlations for calibration measures between the two workouts, regardless of experience group. Nonetheless, experienced athletes made predictions that correlated between workouts, indicating they had a general awareness of their HIFME-related fitness. This can be useful in developing training and competition programmes, as well as setting appropriate goals, because athletes can use a consistent starting point from which they can progress. Finally, the absence of significant gender differences indicates that more research is required to examine whether HIFME athletes of different genders exhibit similar bias, or whether a larger sample size will lead to the expected patterns of higher male overconfidence.

The present study was not without limitations. A common issue identified throughout the discussion was low sample size. Though I anticipated for a sample of approximately sixty athletes to be sufficiently powered for the relationships between demographic factors and calibration to be visible, this was not the case. Since the effect sizes of some of these relationships tend to be small (e.g. Kolovelonis, 2019; Liverakos et al., 2018), then larger sample sizes such as the ones used in studies in Chapter 2 are necessary.[4] Furthermore, some participants did not finish both workouts or were outliers, which further reduced the sample size. Future studies examining HIFME calibration and demographic factors should thus aim

[4] I should note that I did not have available data for effect size comparisons for similar analyses prior to conducting Study 3. Furthermore, there was low demand from participants who wanted to participate in the study. Areas with a higher population than St Andrews would have likely allowed me to recruit more participants.

to collect data from large HIFME competitions (e.g. the CrossFit Open mentioned in Section 3.1) in the same manner that data are typically collected from running competitions.

A second limitation was related to the varied nature of HIFME. HIFME workouts typically include a wide range of movements. This means that only two workouts cannot represent the entire spectrum of HIFME sufficiently. For example, I did not include weightlifting exercises in the workouts used, despite them being very common in HIFME. This was done to ensure that all participants could complete the prescribed exercises, because weightlifting exercises, such as barbell cleans and snatches, are technically demanding. It would have thus not been possible for non-HIFME participants to learn how to execute them properly and safely during one study session. On a related note, rowing in the Rounds workout could have contributed to the observed lack of experience influence on calibration. Rowing is a complex exercise and a few participants in the non-HIFME group indicated they had been members of rowing teams before. This means that high rowing experience in the non-HIFME group could have reduced the experience difference between the two groups, leading to the non-significant finding. Overall, the complexity and plurality of HIFME workouts and movements renders it difficult to make strong inferences about HIFME calibration using just two workouts. Research that aims to gain a thorough understanding of the factors that contribute to calibration in this exercise modality needs to collect data from a wide range of HIFME activities.

Finally, my presence as the researcher during data collection could have influenced how participants made performance predictions. Calibration studies typically observe patterns of overconfidence (e.g. Kolovelonis, 2019; Kolovelonis & Goudas, 2018), but the present study found no bias for the AMRAP workout, and a tendency towards underconfidence for the Rounds workout. This could result from participants wishing to avoid embarrassment by providing me with low and cautious predictions. A way to control for this issue would be for me to merely provide participants with a questionnaire asking for predictions, and leave the room for the duration of this process, thereby limiting social effects on predictions. Even so, it should be noted that the present way of collecting data is possibly higher in ecological validity than the one proposed, as athletes in HIFME settings are likely to make predictions or assess their performance capacity in the presence of others. Therefore, excluding social factors from this process could reduce its realism.

3.4 CONCLUSION

Chapter 3 investigated the influence of expertise, experience, and gender on HIFME calibration. Overall, results illustrated that, though expertise, gender, and most experience factors were not associated with HIFME calibration, having a HIFME background was an important contributor to metacognitive awareness of performance, and that new athletes should thus aim to increase their familiarity with the relevant exercise modality. They also suggested that, to gain a thorough understanding of calibration in HIFME, we need to examine a wide range of workout formats and exercises, and collect data from large sample sizes. This is important to achieve, because understanding which demographic factors contribute to HIFME calibration and how, can assist athletes with managing their performance, motivation, and injury risk. Nonetheless, both Chapters 2 and 3 demonstrate that demographic factors alone are not sufficient to fully explain exercise calibration. This is because the relationships between each factor and calibration are not always consistent, and can have small effect sizes. It is thus imperative to complement literature on demographic factors with research that explores the influence of non-demographic traits on calibration. This is the aim of Chapter 4, in which I explore whether self-reports of exercise metacognition and cognitive calibration can inform us about exercise calibration.

CHAPTER 4: ASSOCIATIONS OF METACOGNITION SELF-REPORTS AND COGNITIVE CALIBRATION WITH EXERCISE CALIBRATION

4.1 INTRODUCTION & RATIONALE

In Chapters 2 and 3, I investigated how demographic factors influence calibration in running and HIFME. Results showed that expertise, experience, age, and gender can—at least partly—inform us about calibration in different exercise modalities. We should thus take demographic information in consideration when assessing an athlete's metacognitive awareness of their performance. However, these results also demonstrated that we need to account for numerous demographic factors when assessing exercise calibration, as the relationships between individual demographic factors and calibration were often inconsistent and/or small. Furthermore, as discussed in Section 1.3.2, demographic factors can be difficult and time-consuming (e.g. years of running experience) or impossible (e.g. gender and age) to manipulate. That is not to say that we cannot use demographic factors to guide calibration interventions, but rather that calibration interventions should not aim to alter them directly. In contrast, metacognition is malleable (e.g. Gutierrez & Schraw, 2015; Gutierrez de Blume, 2017; Nietfeld et al., 2006), and can thus be manipulated directly, suggesting that it is of interest to investigate its association with calibration. Since calibration is an online measure of metacognitive monitoring accuracy (see Section 1.2.3.2), it becomes important to understand its relationship with other offline (e.g. self-reports) and online (e.g. calibration from other domains) measures of metacognition, and to examine whether interventions targeting metacognitive processes facilitate calibration. In the present chapter, I explored the extent to which we can use self-reports of exercise metacognition and cognitive calibration data to predict HIFME and running calibration.

Research on the relationship between self-reports of metacognition and cognitive calibration has been limited. The studies reviewed in Section 1.3.2.1 reported equivocal findings on the extent to which self-report scores of general cognitive metacognition reflect calibration. On the one hand, most examinations have found no association between offline measures of metacognition and calibration (e.g. Jacobse & Harskamp, 2012; Saraç & Karakelle, 2012; Schraw & Dennison, 1994). On the other hand, some studies have exhibited an often weak link between higher scores in metacognition self-reports and better calibration (e.g. Jang et al., 2020; Schraw, 1997; Tobias et al., 1999). Despite the reviewed literature suggesting that we cannot use self-reports to draw strong inferences about cognitive calibration, we might still use them in conjunction with other variables (e.g. demographic factors) to assist us in diagnosing calibration patterns. Furthermore, to my knowledge, Nietfeld (2003) has been the only researcher to examine metacognition self-reports and athletic performance monitoring—though not calibration—and he found that runners who scored higher on running metacognition were more accurate at monitoring their running performance than runners with lower scores (see Section 1.3.2.1). It is thus of interest to further explore the relationship between metacognition self-reports and exercise calibration to better understand whether we can use metacognition inventories to also assess an athlete's calibration. Given the cost-effectiveness and time-efficiency of offline measures of metacognition (Harrison & Vallin, 2018; Schellings & Hout-Wolters, 2011; Veenman, 2011; Winne & Perry, 2000), a potential association with online measures would assist coaches, instructors, and athletes in assessing athlete calibration.

Chapter 4 also explored whether we can use calibration in a cognitive task to predict calibration in an exercise task. There have been numerous suggestions that metacognition contains several domain-general processes (Arbuzova et al., 2020; Carpenter et al., 2019; MacIntyre et al., 2014; Mazancieux, Fleming, Souchay, & Moulin, 2020; Morales, Lau, & Fleming, 2018). Based on such suggestions, we would expect a student who engages in metacognitive processes in cognitive tasks, such as predicting, planning, monitoring, and evaluating academic performance, to engage in similar metacognitive behaviour in physical activity. Accordingly, previous research has demonstrated that self-regulation skills can transfer across domains in young athletes (Jonker et al., 2010, 2011; Jonker, Gemser, & Visscher, 2009; Mccardle, 2015). Jonker and colleagues (2010) found that 128 12-16-year-old elite youth football players reported a higher general tendency to use self-regulation skills, such as self-monitoring and planning, than 164 non-athlete, age-matched controls. Athletes were also more likely to enrol in the pre-university academic system than controls, indicating higher academic achievement. These findings suggest that sports participation facilitates the development of self-regulatory and metacognitive skills, which also transfer to the academic domain. Metacognitive processes that contribute to calibration (e.g. performance monitoring and evaluation) are thus also likely to be present across domains. This means that calibration in academic tasks could correlate with calibration in exercise tasks.

Though not comparing cognitive and exercise calibration directly, Arbuzova and colleagues (2020) examined domain-generality of metacognition by comparing metacognitive efficiency (i.e. whether participants had high confidence for correct responses and low confidence for incorrect responses) across motor, visuomotor, and visual conditions of a computerised ball throwing task. The conditions differed in the extent to which participants received only visual, only motor, or both visual and motor feedback for their ball throwing performance before they selected the correct trial outcome out of two available options. For each trial, participants were required to rate their confidence regarding the accuracy of their response. The results exhibited moderate positive correlations for metacognitive efficiency between conditions, providing evidence for the domain-generality of feedback monitoring processes. Though these findings suggest that metacognition and calibration could also generalise across cognitive and exercise domains, they do not provide direct evidence for it. The motor and visuomotor conditions used by Arbuzova and colleagues differed to most athletic tasks, as they were simple and did not require physical exertion. It is thus important to examine whether calibration in cognitive tasks correlates with calibration in athletic tasks directly. Only then will it be possible to make strong inferences on calibration generalisation across cognitive and exercise domains.

Overall, the purpose of Chapter 4 was to examine the extent to which we can use self-reports of exercise metacognition and calibration for cognitive tasks to make inferences about an athlete's exercise calibration. To explore these relationships, I presented and analysed data from three different studies. In Studies 4 and 5, I investigated whether metacognition self-reports would predict calibration in HIFME and running. Based on previous evidence, I anticipated that, if self-reports of metacognition did predict exercise calibration, the effect size of this relationship would likely be small or small-to-moderate. In Study 6, I tested whether calibration in a memory recognition task would be associated with calibration in the HIFME tasks used in Study 3 presented in Chapter 3. Following previous self-report and motor calibration research, I expected that bias and absolute accuracy in the HIFME workouts would correlate with bias and absolute accuracy in the prospective and retrospective metacognitive judgments made by participants for the memory recognition task.

4.2 STUDY 4 – SELF-REPORTS & HIFME CALIBRATION

4.2.1 Study specifics

The goal of Study 4 was to examine whether self-reports of general exercise metacognition would contribute to calibration in an unfamiliar exercise task. To collect self-report data of exercise metacognition, I developed the Metacognitive Awareness Inventory for Exercise (MAIE; see Section 4.2.2.2) by adapting the Metacognitive Awareness Inventory (MAI; see Section 4.2.2.2; Schraw & Dennison, 1994). I measured calibration using a HIFME task, which I selected based on its complexity and likely novelty to participants. I wanted to examine the relationship between self-reports of metacognition and calibration in exercise using an unfamiliar task, as it would be interesting to know whether self-perceptions of general exercise metacognition are associated with how participants make predictions in an exercise modality they have not experienced before. To ensure task novelty for all participants, I only recruited athletes without previous HIFME experience. Given suggestions from cognitive research that self-report inventories that are not specific to the tasks used to collect online metacognition data can reduce the likelihood of observing a relationship between offline and online measures of metacognition (Schellings, 2011; Schellings et al., 2013), I was uncertain as to whether MAIE scores would predict HIFME calibration. As part of a secondary examination, I used the MAI to collect data on academic metacognition self-reports to test for domain-generality of metacognition self-reports across exercise and cognition. I anticipated that MAI components would exhibit a positive correlation with MAIE components, in line with research suggesting that metacognitive skills transfer across the academic and exercise domains.

4.2.2 Methods

4.2.2.1 Participants

Participants were recreational athletes who reported no previous HIFME experience, i.e. had either never engaged in HIFME or their cumulative HIFME experience was less than a month (meaning that even if they had tried HIFME before, they were still eligible to participate if they had not engaged in it consistently—e.g. for more than a total of ten hours). The criteria for HIFME and non-HIFME experience were the same as those described in Section 3.2.2.1.

I recruited 54 participants between the ages of 18 and 40 years old from the student and local population of St Andrews. Four participants could not finish the HIFME workout or misunderstood workout instructions (e.g. one participant thought the workout prescribed 10 lunges instead of 20; see Section 4.2.2.3 for workout description), five reported previous HIFME experience, and one was an outlier (absolute z score for absolute accuracy percentage higher than three), so their data were excluded from the calibration analyses. After exclusions, I used 54 (37 women and 17 men; M_{age} = 20.6 years old, *SD* = 2.6 years old) data-points for metacognition self-report comparisons (no exclusion based on experience, workout completion, or calibration), and 44 (29 women and 15 men; M_{age} = 20.3 years old, *SD* = 2.2 years old) for calibration analyses.

Ethical approval was granted from the University of St Andrews School of Psychology & Neuroscience Ethics Committee (Ethics approval code: PS13905; see Appendix 8.1.4). All participants were compensated at a rate of £5/hour.

4.2.2.2 Questionnaires

Metacognitive Awareness Inventory (MAI; see Appendix 8.7). The MAI assesses the use of metacognition in learning, and consists of 52 items (Schraw & Dennison, 1994). For each statement, participants indicated whether they agreed on a fully labelled scale comprising of 1 (strongly disagree), 2 (sometimes disagree), 3 (neutral), 4 (sometimes agree), and 5 (strongly agree). Based on the suggestions by Harrison and Vallin (2018), I calculated scores for knowledge of cognition (17 questions; e.g., "I know what kind of information is most important to learn") and regulation of cognition (35 questions; e.g., "I ask myself questions about the material before I begin"). Knowledge of cognition refers to declarative (i.e. knowledge of one's academic skills and abilities), procedural (i.e. knowledge on how to apply metacognitive processes to learning), and conditional knowledge (i.e. knowing why and when to apply metacognitive skills and strategies to achieve effective learning). Regulation of cognition in MAI refers to planning (i.e. setting goals and allocating resources prior to learning), information management strategies (i.e. skills and strategies used to process information more efficiently), comprehension monitoring (i.e. monitoring understanding of material), debugging strategies (i.e. strategies to correct comprehension and performance errors), and evaluation (i.e. analysing and evaluating performance and strategy effectiveness). Higher scores indicate higher academic metacognition.

Metacognitive Awareness Inventory for Exercise (MAIE; see Appendix 8.8): I developed MAIE based on the structure and questions of MAI to measure metacognition in exercise and sports settings. The MAIE consists of 50 items—two fewer than the MAI, as I found it difficult to adapt two of the items to an exercise setting accurately. For each statement, participants indicated whether they agreed on a fully labelled scale comprising of 1 (strongly disagree), 2 (sometimes disagree), 3 (neutral), 4 (sometimes agree), and 5 (strongly agree). I calculated scores for knowledge of exercise performance (17 questions; e.g., "I perform best when I have experience with the sport/exercise") and regulation of exercise (33 questions; e.g., "I set specific goals before I begin a workout"). The knowledge component refers to declarative, procedural, and conditional knowledge of exercise performance (same as the MAI knowledge component). The regulation component refers to performance planning, exercise management strategies, performance monitoring, debugging strategies, and exercise performance evaluation. Higher scores indicate higher exercise metacognition.

4.2.2.3 HIFME workout

The HIFME workout consisted of participants completing as many repetitions of the prescribed exercises as possible within ten minutes. The format was similar to the AMRAP workout in Chapter 3 (see Section 3.2.2.2). For each round, I asked participants to first complete 10 burpees (participants had to start from a standing position, get their chest on the floor, stand back up, and jump and clap at the top; Fig. 4.1), then 10 sit-ups (using a sit-up "pillow" placed behinds participants' back; Fig. 4.2), and, finally, 20 alternating lunges (i.e. athletes had to alternate legs for every repetition – 10 lunges for each leg; Fig. 4.3). Participants could only proceed to the next exercise after completing all repetitions for the previous exercise. I measured performance in terms of the total number of repetitions completed in 10 minutes; e.g., five rounds and 10 repetitions were equal to 210 repetitions. Participants made their predictions in terms of the number of repetitions they expected to complete.

Figure 4.1

Demonstration of burpees.

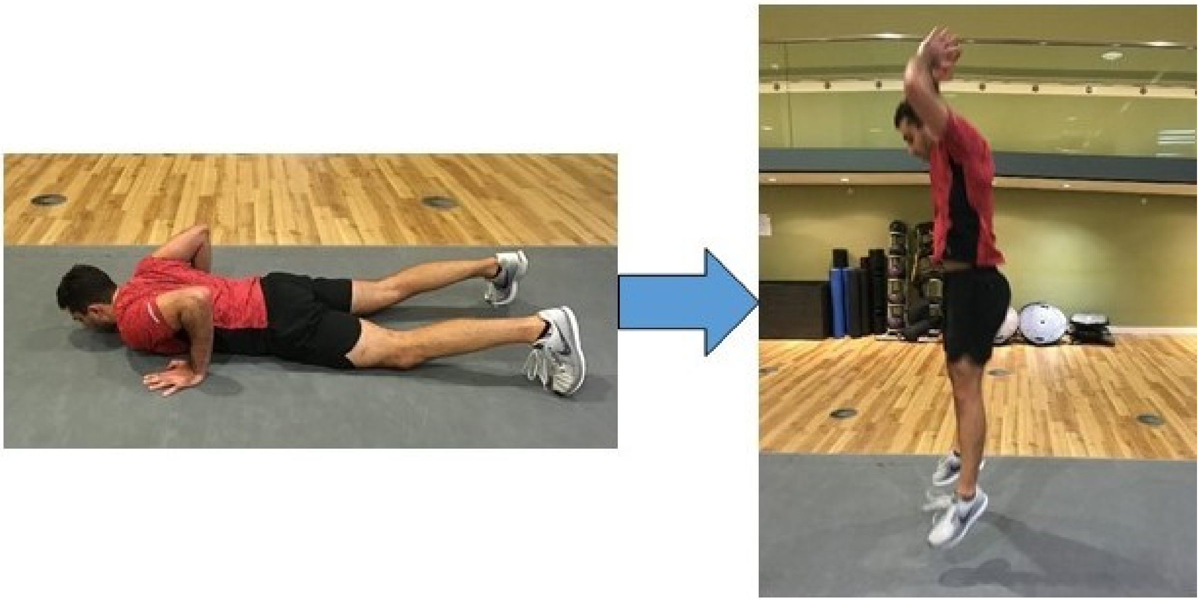


Figure 4.2

Demonstration of sit-ups.

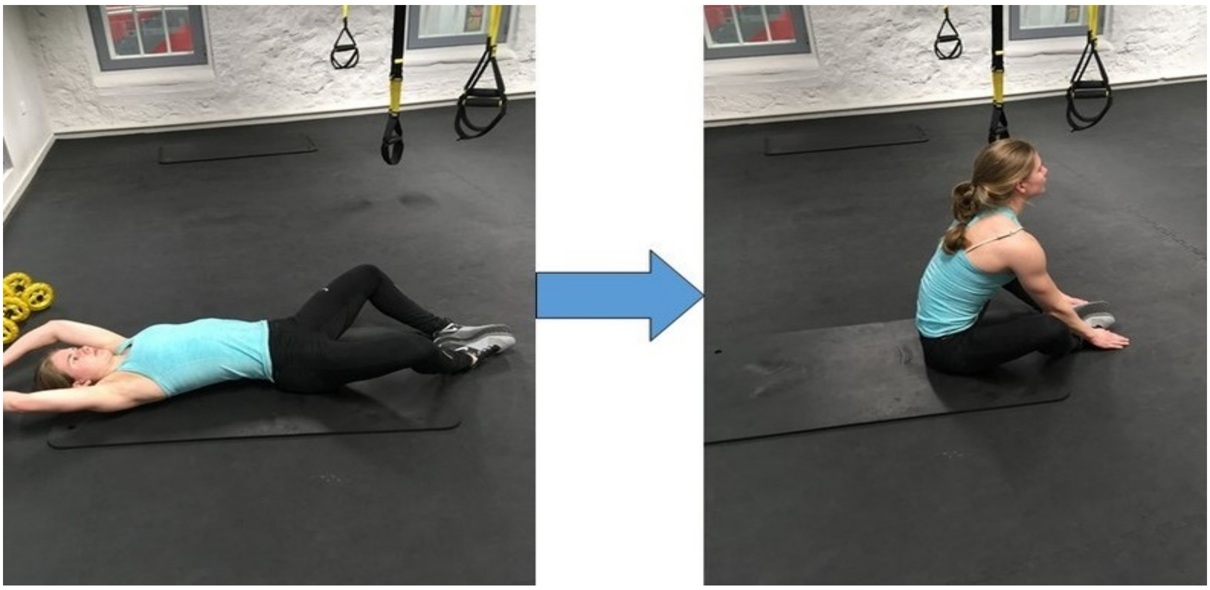
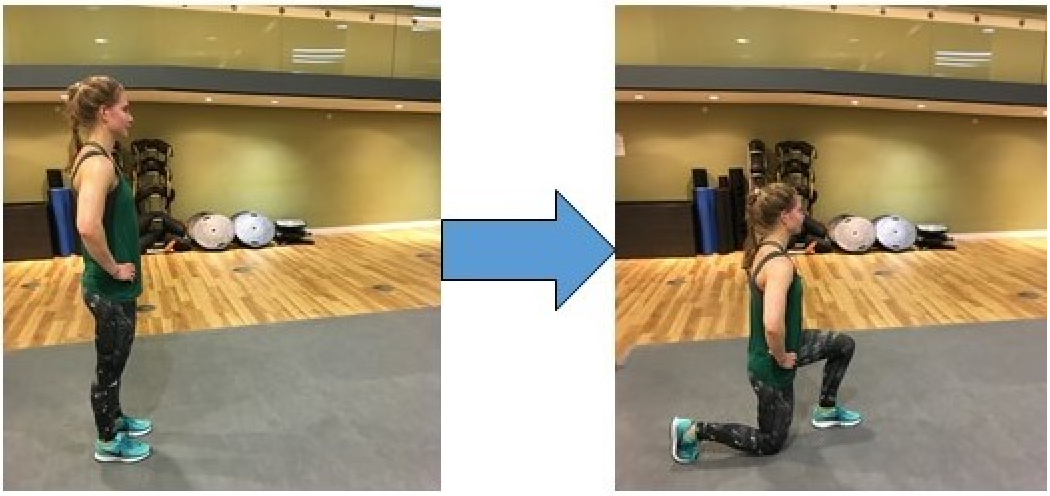


Figure 4.3

Demonstration of lunges.



4.2.2.4 Design

My primary analysis was on the examination of the associations between MAIE components (knowledge and regulation) and HIFME calibration. I assessed calibration using bias and absolute accuracy percentages relative to performance (formulae for these calculations in HIFME can be found in Section 3.2.2.4). Using percentages allowed me to control for the effects of performance value variation on calibration. In my secondary analysis, I examined the domain-generality of metacognition self-reports by conducting correlations between MAIE and MAI components.

4.2.2.5 Procedure

I emailed participants with a Participant Information Sheet and a PAR-Q health eligibility form (see Appendix 8.5.2) to ensure that no previous health conditions would be exacerbated by the HIFME workout. During the testing session, I asked participants to provide their informed consent, verbally describe their previous exercise experience, and complete the MAI and the MAIE. The order of the metacognitive questionnaires was counterbalanced. During a 10-minute warm-up period, I demonstrated the standards and technique for each exercise before the participants performed them themselves. I then presented participants with the HIFME workout, and I explicitly asked them to provide realistic predictions (i.e. "how [they] thought [they] were going to perform"), as opposed to goal predictions (i.e. "how [they] hoped [they] were going to perform"). I also prompted them to be as specific as possible in their predictions (i.e. try to provide a prediction in terms of rounds and repetitions, rather than just rounds). After completing the workout, I debriefed participants, and the session concluded.

4.2.3 Results

4.2.3.1 Metacognition self-reports

Internal Reliability. MAIE Knowledge (α = .82) and Regulation (α = .93), and MAI Knowledge (α = .80) and Regulation (α = .86) had good internal reliability.

Metacognitive Knowledge and Regulation. I conducted correlations between knowledge and regulation components for MAIE and MAI (see Table 4.1 for correlation coefficients). The knowledge and regulation components exhibited a moderate-to-strong positive correlation with each other in both MAIE and MAI, indicating that higher metacognitive knowledge of cognition and exercise performance are associated with higher regulation in their respective modalities.

Table 4.1

Table illustrating correlations between the knowledge and regulation components of MAIE and MAI.

	MAIE_Know	MAIE_Reg	MAI_Know	MAI_Reg
MAIE_Know	—	.70**	.64**	.60**
MAIE_Reg		—	.50**	.63**
MAI_Know			—	.66**
MAI_Reg				—

Note. MAIE_Know And MAIE_Reg represent metacognitive knowledge and regulation in exercise respectively. MAI_Know and MAI_Reg represent metacognitive knowledge and regulation in learning respectively. *. Significant at the .05 level (2-tailed). **. Significant at the .01 level (2-tailed). I have adjusted significance levels using the Holm correction.

Domain generality of self-reports. I conducted correlations across MAIE and MAI components to examine domain generality of metacognition self-reports (Table 4.1). All components significantly correlated with each other, even after using the Holm correction for multiple comparisons. MAIE knowledge exhibited a moderate-to-strong positive correlation with MAI knowledge and MAI regulation. MAIE regulation exhibited a moderate-to-strong positive correlation with MAI regulation and a moderate positive correlation with MAI knowledge. These results support domain generality of metacognition self-reports across exercise and academic modalities. However, there was variance that was unique to MAIE and MAI components, also providing evidence for domain specificity. Furthermore, each component (e.g. regulation) exhibited smaller correlations with the incongruent component of the other questionnaire (e.g. knowledge) than the congruent one, suggesting that congruent components were more strongly associated with each other than they were with their incongruent components.

4.2.3.2 Performance & Predictions

To examine the associations of MAIE knowledge and regulation with performance and predictions, I conducted multiple regressions where I entered both factors in the model at the same time. I did this because the two components were strongly correlated, so I was interested in the extent to which the individual association of each component (which I investigated using bivariate correlations) with the outcome variable would change after accounting for variance from the other component. At the same time, the two components were not so highly correlated as to have multicollinearity in the regression model (i.e. $r > .80$). I implemented the same method of analysis in Study 5 in Section 4.3.3.2.

Performance. Results from correlational analyses and the multiple regression model for HIFME performance can be seen in Tables 4.2 and 4.3. The regression model was significant in explaining performance variance, $R^2 = .218$, $F_{(2, 41)} = 5.72$, $p = .006$. MAIE regulation scores significantly predicted the number of repetitions completed, with participants who self-reported as regulating their athletic performance more being better performers than participants who reported lower exercise regulation. After accounting for MAIE regulation variance, MAIE knowledge did not exhibit a significant association with performance. This contrasted correlation results, where higher scores in both components we positively associated with better performance.

Table 4.2

Correlation coefficients for the individual associations of MAIE components with the outcome variables performance and predictions.

	Performance		**Predictions**	
Factor	***r***	***p***	***r***	***p***
MAIE_Know	0.32	.034	0.17	.280
MAIE_Reg	0.46	.001	0.28	.064

Note. *r* represents the correlation coefficient of each factor with the outcome variables. *p* represents the *p*-value associated with each corresponding factor and outcome variable. MAIE_Know represents the knowledge component of MAIE, and MAIE_Reg represents the regulation component.

Predictions. Results from correlational analyses and the multiple regression model for HIFME predictions can be seen in Tables 4.2 and 4.3. The regression model was not significant in explaining HIFME prediction variance, $R^2 = .085$, $F_{(2, 41)} = 1.91$, $p = .161$. MAIE regulation scores only showed a non-significant tendency to predict HIFME predictions, with higher scores predicting higher predictions, whereas MAIE knowledge score did not exhibit an association with predictions. Regressions results were in line with correlation findings. Overall, neither MAIE component was a significant contributor to HIFME predictions.

Table 4.3

Multiple regression coefficients for MAIE component predictors on the outcome variables predictions and performance.

Outcome variable	Coefficient	*B*	*Beta*	*Std. Error*	*t*	*p*
Performance	Intercept	86.87	—	52.47	1.66	.105
	MAIE_Know	-0.42	-0.08	1.16	-0.37	.717
	MAIE_Reg	1.11	0.52	0.45	2.46	.018
Predictions	Intercept	120.62	—	109.03	1.11	.275
	MAIE_Know	-1.21	-0.12	2.41	-0.50	.619
	MAIE_Reg	1.51	0.37	0.94	1.61	.116

Note. Factors were entered at the same time in each multiple regression model. B and Beta represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. MAIE_Know is Metacognitive knowledge in exercise, and MAIE_Reg is Metacognitive regulation in exercise.

4.2.3.3 Calibration

Similar to performance and prediction analyses, I conducted correlational analyses and multiple linear regressions to examine whether MAIE components contribute to bias and absolute accuracy. In each regression model, I entered all predictors at the same time to account for shared variance. For each outcome variable, the first regression model did not include performance, whereas the second model did. I did this because performance was associated with the regulation component in the previous analyses (see Section 4.2.3.2), so it could influence its relationship with calibration. I was thus interested in the extent to which regulation's association with each calibration measure would change with the inclusion of performance in the model.

Bias. In Study 4, positive bias scores indicate overconfidence, and negative bias scores indicate underconfidence. Participants in the sample were significantly overconfident in their predictions (*M* = 14.87%, *SD* = 37.69%), $t_{(43)} = 2.62$, $p = .012$, $d = 0.39$.

Table 4.4

Correlation coefficients for the individual associations of MAIE components and performance with the outcome variables HIFME bias and absolute accuracy.

	Bias		**Abs Acc**	
Factor	***r***	***p***	***r***	***p***
MAIE_Know	0.02	.921	-0.01	.949
MAIE_Reg	0.05	.766	-0.03	.831
Performance	-0.04	.799	-0.11	.492

Note. *r* represents the correlation coefficient of each factor with the outcome variables. *p* represents the *p*-value associated with each corresponding factor and outcome variable. MAIE_Know represents the knowledge component of MAIE, and MAIE_Reg represents the regulation component.

Results from correlational analyses and the multiple regression models for bias can be seen in Tables 4.4 and 4.5. Neither regression model was significant in predicting bias, regardless of whether they included performance, $R^2 = .008$, $F_{(3, 40)} = 0.11$, $p = .955$, or not, $R^2 = .003$, $F_{(2, 41)} = 0.06$, $p = .939$. MAIE components (Fig. 4.4A & 4.4C) and performance (Fig. 4.4E) all failed to contribute to HIFME bias, mirroring findings from correlations between each factor and bias.

Absolute Accuracy. Results from correlational analyses and the multiple regression models for absolute accuracy can be seen in Tables 4.4 and 4.5. As with bias, neither regression model was significant in predicting HIFME absolute accuracy, regardless of whether they included performance, $R^2 = .012$, $F_{(3, 40)} = 0.16$, $p = .922$, or not, $R^2 = .002$, $F_{(2, 41)} = 0.03$, $p = .967$. Participant precision was not influenced by either MAIE component (Fig. 4.4B & 4.4D), or performance (Fig. 4.4F). Overall, both correlational analyses and multiple regression models showed that MAIE components and HIFME performance did not contribute to HIFME bias or absolute accuracy.

Table 4.5

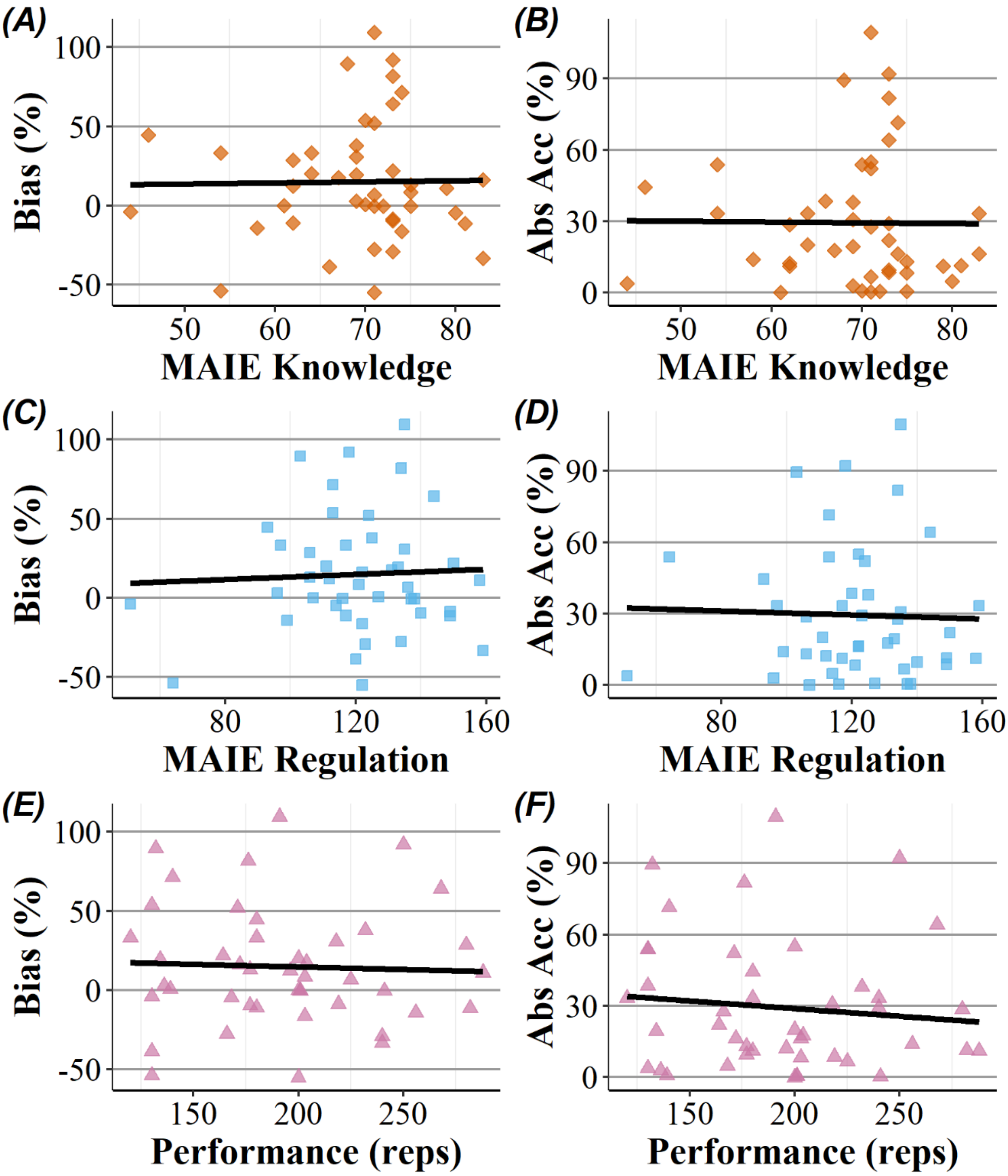
Multiple regression coefficients for MAIE component and performance predictors on the outcome variables bias and absolute accuracy.

Outcome variable	**Performance inclusion**	**Coefficient**	***B***	***Beta***	***Std. Error***	***t***	***p***
Bias	No Performance	Intercept	12.02	—	48.72	0.25	.806
		MAIE_Know	-0.21	-0.05	1.08	-0.20	.846
		MAIE_Reg	0.14	0.08	0.42	0.34	.735
	Performance included	Intercept	17.73	—	50.82	0.35	.729
		MAIE_Know	-0.24	-0.05	1.09	-0.22	.828
		MAIE_Reg	0.22	0.12	0.46	0.48	.637
		Performance	-0.07	-0.08	0.15	-0.45	.656
Absolute Accuracy	No Performance	Intercept	30.55	—	35.78	0.85	.398
		MAIE_Know	0.12	0.04	0.79	0.15	.880
		MAIE_Reg	-0.08	-0.06	0.31	-0.25	.802
	Performance included	Intercept	36.57	—	37.23	0.98	.332
		MAIE_Know	0.09	0.03	0.80	0.11	.910
		MAIE_Reg	0.00	0.00	0.33	0.00	.998
		Performance	-0.07	-0.12	0.11	-0.65	.522

Note. Factors were entered at the same time in each multiple regression model. B and Beta represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. MAIE_Know is Metacognitive knowledge in exercise, and MAIE_Reg is Metacognitive regulation in exercise.

Figure 4.4

Scatterplots illustrating the relationships between MAIE components, performance, and calibration bias and absolute accuracy.



Note. Panels A and C show the influence of MAIE Knowledge and Regulation on bias. Panels B and D show the influence of MAIE Knowledge and Regulation on absolute accuracy. Panels E and F show the contributions of HIFME performance on bias and absolute accuracy. Reps stands for repetitions completed in the HIFME workout.

4.2.4 Discussion of Study 4

The purpose of Study 4 was to examine the relationship between general exercise metacognition self-reports and calibration in an unfamiliar HIFME workout. Additionally, I was interested in the extent to which exercise and academic metacognition self-reports would correlate with each other, indicating domain-generality of metacognition across modalities.

Given the equivocality of previous findings in cognition, I made no specific prediction as to whether self-reports of general exercise metacognition would predict calibration in an unfamiliar HIFME task. I found no relationship between self-reports of exercise metacognition and HIFME calibration, as MAIE knowledge and regulation scores were not predictive of bias and precision in a novel exercise task. This finding was not surprising, as it was in line with previous cognitive research that did not observe a connection between metacognition self-reports and calibration (Schraw & Dennison, 1994; Sperling et al., 2004), and suggestions that metacognition questionnaires should be task-specific in order to correlate with online measures of metacognition (Schellings, 2011; Schellings et al., 2013).

Nonetheless, results from Study 4 are not necessarily incompatible with Nietfeld's (2003) finding of participants with higher self-reports of running metacognition exhibiting better running monitoring accuracy, and do not illustrate definitively that metacognition questionnaires cannot inform us about exercise calibration. The MAIE used reflected general exercise metacognition, so it was likely not specific enough to predict HIFME calibration. Furthermore, though domain generality of metacognition would suggest that athletes with high metacognitive skills should be expected to implement these skills in a new activity, participants in the study did not have any previous experience with HIFME, and thus had not engaged in HIFME metacognition before. Thus, even individuals who reported frequent engagement with metacognitive behaviour might not have been able to use their metacognitive skills successfully to inform their predictions for an unfamiliar task. Perhaps experience with an activity is a prerequisite for metacognitive self-reports to predict calibration in the examined activity. Indeed, there have been previous suggestions regarding the presence of an interaction between self-reports of metacognition and experience in relation to cognitive calibration (Jang et al., 2020). Because of the aforementioned possibilities, it is important to examine the relationship between metacognition self-reports and exercise calibration using a task-specific questionnaire and a familiar exercise modality.

In accordance with my expectations of domain-generality of metacognition across cognition and exercise, participants who scored highly on MAIE components were also more likely to score highly on MAI. Interestingly, the scores between the two inventories were not identical, suggesting that both domain-general and domain-specific aspects of metacognition contribute to self-report scores. Though the positive association of self-report scores across the two domains was in line with previous research on the generalisability of self-regulation and metacognition self-report scores (Jonker et al., 2010, 2011; Mccardle, 2015), the validity of the present result is not clear. Since I developed MAIE using MAI, it is likely that similarities in the structure and items between the two questionnaires contributed to the observed association. Thus, we need to compare scores in questionnaires that have been developed independently to further examine the reliability and validity of the present finding.

4.3 STUDY 5 – SELF-REPORTS & RUNNING CALIBRATION

4.3.1 Study specifics

The purpose of Study 5 was to examine the relationship between metacognition self-reports and exercise calibration using a running-specific questionnaire and a running calibration task. As discussed above, using task-specific metacognition questionnaires and collecting calibration data from familiar tasks could result in a more robust relationship between offline and online measures of metacognition, leading to less equivocal results. I thus anticipated that, contrary to Study 4, where MAIE collected general exercise metacognition data and participants did not have previous HIFME experience, participants with previous running experience in Study 5 would report running metacognition scores that would predict their running calibration. To assess running-specific self-reports of metacognition, I developed the Metacognitive Awareness Inventory for Running (MAIR; see Section 4.3.2.2), which was based on the MAIE developed in Study 4. To measure running calibration, I asked participants to produce impulsive and strategic predictions for a 1km running trial (see Section 4.3.2.3 for task details and Section 4.3.2.4 for prediction instructions). I collected different prediction types as part of a prediction strategy manipulation, which I discuss further in Study 8 in Chapter 5. Comparisons between the two prediction types are thus outside the scope of Chapter 4.

4.3.2 Methods

4.3.2.1 Participants

All participants needed to have had at least one year of running experience, and to typically run at least twice per week during the period leading to the study. For the purposes of the study, participants only had running experience when it was not part of another sport, e.g. participants who only ran whilst playing football were not eligible to participate. For participants to be eligible to participate, they had to engage in some form of independent running (e.g. going out for runs or training for a running race).

Sixty-seven runners (33 men and 34 women; M_{age} = 23.3 years old, *SD* = 5.5 years old) between the ages of 18 and 40 years old participated in the study, two of which did not finish the workout. Overall, 63 runners (30 men and 33 women; M_{age} = 23.3 years old, *SD* = 5.6 years old) completed both the running workout and the metacognition self-reports (two participants selected multiple answers instead of one, so their answers were invalid). I excluded three runners from the impulsive prediction analyses, as they indicated they gave strategic, instead of impulsive, predictions (see Section 4.3.2.4). Furthermore, there were two outliers (i.e. the absolute value of *z* scores for absolute accuracy percentages was larger than three), leading to a final sample of 58 runners (28 men and 30 women; M_{age} = 23.3 years old, *SD* = 5.4 years old) for the impulsive prediction calibration analyses. There were no outliers for strategic predictions, so I used the sample of all 63 participants who completed the running workout and the questionnaires for the strategic prediction calibration analyses.

Ethical approval was granted from the University of St Andrews School of Psychology & Neuroscience Ethics Committee (Ethics approval code: PS14429; see Appendix 8.1.5). All participants were compensated at a rate of £5/hour.

4.3.2.2 Questionnaire

Metacognitive Awareness Inventory for Running (MAIR; see Appendix 8.9). MAIR is a 50-item questionnaire that assesses self-reported use of metacognition in running. I developed MAIR by adapting the MAIE I developed in Study 4. Eighteen items I could not adapt from MAIE to running directly remained the same, as they addressed general exercise metacognition. Participants had to indicate whether each sentence applied to them on a scale from one (Strongly Disagree) to five (Strongly Agree)—all scale choices had corresponding values. MAIR questions assess knowledge about running performance (17 items; e.g., "I know when each running strategy I use will be most effective") and regulation of performance (33 items; e.g., "I ask myself periodically if I am meeting my running goals") in a similar way to MAIE in Section 4.2.2.2. Higher MAIR scores in each component indicate higher running metacognition.

4.3.2.3 Running workout

In the running workout, I asked participants to run one kilometre as quickly as they could. I measured predictions and finish time in terms of the minutes and seconds taken to complete the distance. I chose to examine calibration using the distance of one kilometre assuming that runners participating in the study would be familiar with it, without necessarily knowing exactly what their typical time for it is (as might have been the case with another distance such as one mile). Participants warmed up and completed the running task using an indoor motorised treadmill set at 1% incline to simulate outdoor running oxygen uptake demands (Jones & Doust, 1996). Participants ran at a self-determined pace, which they adjusted using a lever on the right hand side of the treadmill. All information other than distance run was covered and hidden from the participant using a black cover. I did this to further simulate outdoors running where measurements such as speed are often unavailable, and to ensure that participants did not just use the screen information to adjust their pace to match their predictions. To start the workout, each participant had to increase the treadmill speed until they started running. At that point, I started counting the time and the distance covered. As soon as the participant finished the 1km trial, the workout was completed.

4.3.2.4 Prediction instructions

Prior to the workout, I provided participants with instructions on how to make their predictions. First, I informed them they had to make estimates based on how they thought they were going to perform, rather than how they hoped to do so. Additionally, I instructed them to be as specific as possible (i.e. include seconds in their predictions if they thought they would not finish after exactly a number of minutes). Finally, I asked them to wait until they had received instructions on how to make their predictions before they started thinking about them. After I presented them with the workout, I gave them the following instructions for their impulsive predictions: “I want you to give me a prediction based on your gut feeling. Simply give me the first prediction that comes to your mind, without engaging in any strategic thinking.” If a participant tried to engage in strategic thinking, I advised them not to. Following this, I instructed participants to provide their strategic predictions using the following instructions: “I want you to think about the prediction you give me and be strategic about it. Do not just provide me with an impulsive prediction.” Following each prediction, I asked participants to confirm that their prediction was impulsive or strategic. (I provide explanation for the use of impulsive and strategic predictions in Study 8 in Chapter 5).

4.3.2.5 Design

The purpose of Study 5 was to examine the associations between the MAIR and calibration. I thus tested the extent to which scores for each MAIR component could predict impulsive and strategic prediction bias and absolute accuracy (formulae for running bias and absolute accuracy percentage calculations can be found in Section 2.2.2.3). I could not counterbalance the order of prediction instructions, as asking for strategic instructions first would lead to the same predictions across priming conditions.

4.3.2.6 Procedure

I emailed participants with a Participant Information Sheet that contained details about the study, a PAR-Q health eligibility form (see Appendix 8.5.2) that ensured no previous health conditions would be exacerbated during the running workout, and an informed consent form. If eligible, I assigned participants a study slot. During the experimental session, participants first completed a running experience questionnaire (see Appendix 8.10) and the MAIR. Following this, they warmed up by running for five minutes on the treadmill at a self-selected pace in an identical process to the workout. I also gave them the opportunity to stretch or run more if they wished to warm up for longer. I then provided participants with specifications on what their performance predictions would entail, presented them with the running workout, and asked them to predict their performance twice—first, following impulsive instructions, and second, following strategic instructions. After completing the workout, I debriefed participants, and the session concluded.

4.3.3 Results

4.3.3.1 Metacognition

Internal Reliability. MAIR Knowledge (α = .80) and Regulation (α = .88) components had good internal reliability.

Metacognitive Knowledge and Regulation. I conducted a correlation between the MAIR knowledge and regulation components to examine their relationship. Similar to results for MAI and MAIE in Section 4.2.3.1, the knowledge and regulation components exhibited a strong positive correlation between them, $r_{(63)} = 0.71$, $p < .001$. Runners with high metacognitive knowledge scores were more likely to also record high metacognitive regulation scores.

4.3.3.2 Performance & Predictions

I analysed the contributions of MAIR components to running performance and predictions using correlational analyses and multiple linear regression models as described in Section 4.2.3.2.

Performance. Results from correlational analyses and the multiple regression model for performance can be seen in Tables 4.6 and 4.7. The regression model was significant in explaining performance variance, R^2 = .126, $F_{(2, 60)} = 4.31$, p = .018. In line with the correlational analysis, higher MAIR knowledge scores predicted better performance than lower MAIR knowledge scores. Interestingly, there was a non-significant trend for higher MAIR regulation scores to predict worse performance than lower MAIR regulation scores. This was in contrast with the correlational analysis and what we would expect the relationship between metacognitive regulation and performance to be, but it was likely the result of accounting for knowledge variance, with which regulation strongly correlated.

Table 4.6

Correlation coefficients for the individual associations of MAIR components with the outcome variables performance and impulsive and strategic predictions.

	Performance		**Impulsive**		**Strategic**	
Factor	***r***	***p***	***r***	***p***	***r***	***p***
MAIR_Know	-0.27	.035	-0.32	.016	-0.25	.051
MAIR_Reg	-0.02	.878	-0.15	.268	-0.14	.279

Note. *r* represents the correlation coefficient of each factor with the outcome variables. *p* represents the *p*-value associated with each corresponding factor and outcome variable. MAIR_Know represents the knowledge component of MAIR, and MAIR_Reg represents the regulation component.

Predictions. Results from correlational analyses and multiple regression models for impulsive and strategic predictions can be seen in Tables 4.6 and 4.7. The regression models were significant in explaining impulsive prediction variance, $R^2 = .109$, $F_{(2, 55)} = 3.37$, $p = .042$, but not strategic prediction variance, $R^2 = .063$, $F_{(2, 60)} = 2.03$, $p = .140$. Regression results mirrored those of correlational analyses. The MAIR regulation component did not contribute to either type of prediction, whilst MAIR knowledge was a significant predictor of impulsive predictions, and a non-significant predictor of strategic predictions. Overall, there was a tendency for participants who self-reported higher running metacognitive knowledge to predict faster trial completion times, mirroring performance results.

Table 4.7

Multiple regression coefficients for MAIR component predictors on the outcome variables performance and predictions.

Outcome variable	Coefficient	*B*	*Beta*	*Std. Error*	*t*	*p*
Performance	Intercept	338.40	—	43.36	7.81	< .001
	MAIR_Know	-2.80	-0.50	0.95	-2.93	.005
	MAIR_Reg	0.90	0.33	0.46	1.94	.057
Impulsive Predictions	Intercept	445.70	—	77.50	5.75	< .001
	MAIR_Know	-3.82	-0.41	1.65	-2.32	.024
	MAIR_Reg	0.63	0.13	0.83	0.76	.448
Strategic Predictions	Intercept	375.92	—	56.61	6.64	< .001
	MAIR_Know	-2.10	-0.30	1.24	-1.69	.097
	MAIR_Reg	0.24	0.07	0.60	0.40	.694

Note. Predictors were entered at the same time in each multiple regression model. B and Beta represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. MAIR_Know is Metacognitive knowledge in exercise, and MAIR_Reg is Metacognitive regulation in running.

4.3.3.3 Calibration

I analysed the contributions of MAIR components to running bias and absolute accuracy using correlational analyses and multiple linear regression models as described in Section 4.2.3.4.

Bias. In Study 5, positive bias scores indicate underconfidence, and negative bias scores indicate overconfidence. Participants in the sample were significantly underconfident in their impulsive predictions (M = 6.33 %, SD = 23.47%), $t_{(57)} = 2.05$, $p = .045$, $d = 0.27$, and non-significantly underconfident in their strategic predictions (M = 3.52 %, SD = 15.39%), $t_{(62)} = 1.82$, $p = .074$, $d = 0.23$.

Table 4.8

Correlation coefficients for the individual associations of MAIR components and performance with the outcome variables bias and absolute accuracy for impulsive and strategic predictions.

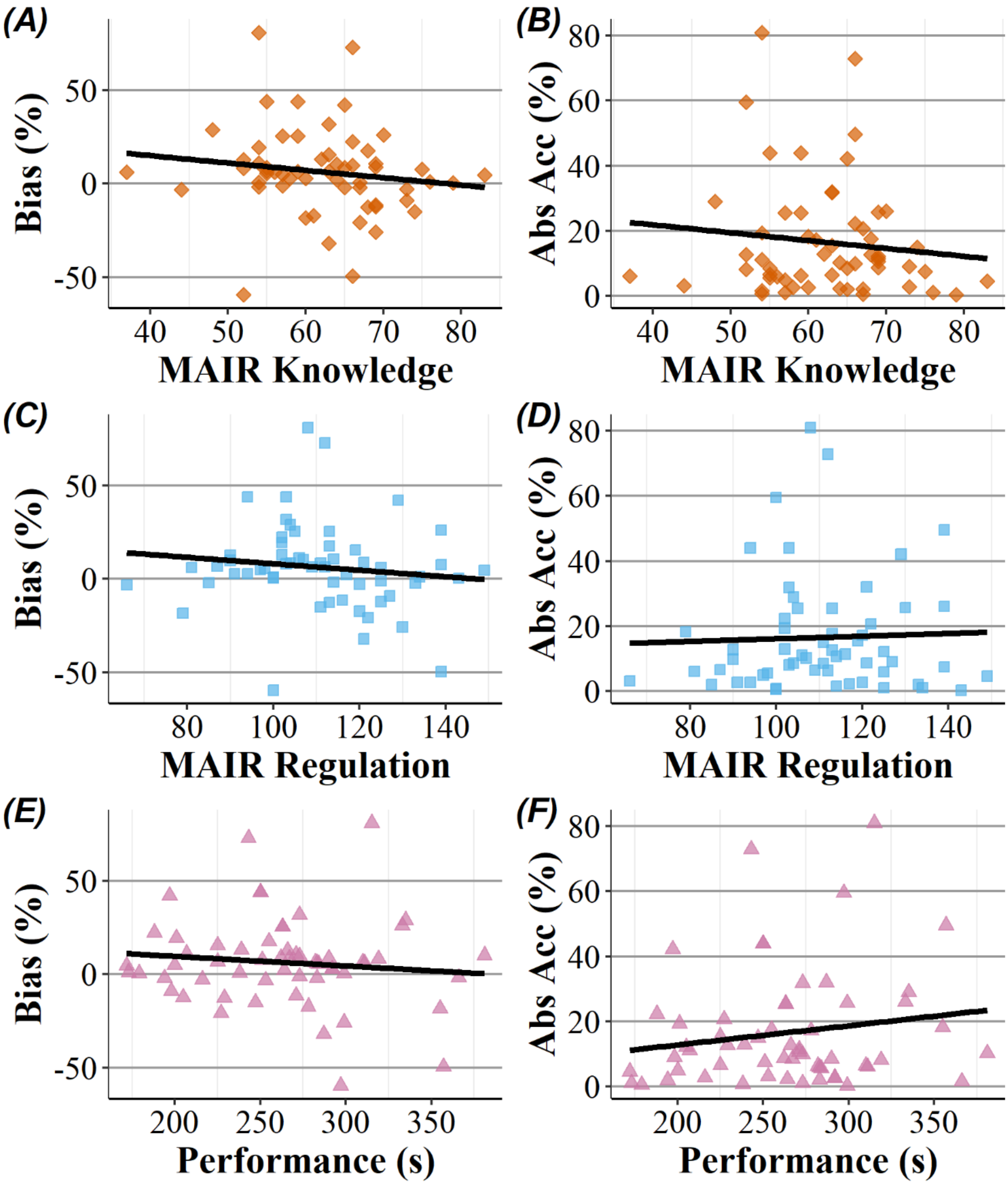
	Impulsive Predictions				**Strategic Predictions**			
	Bias		**Abs Acc**		**Bias**		**Abs Acc**	
Factor	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***
MAIR_Know	-0.14	.284	-0.12	.384	-0.05	.718	0.04	.770
MAIR_Reg	-0.13	.351	0.04	.773	-0.17	.172	-0.02	.862
Performance	-0.11	.421	0.16	.219	-0.19	.140	0.14	.285

Note. *r* represents the correlation coefficient of each factor with the outcome variables. *p* represents the *p*-value associated with each corresponding factor and outcome variable. MAIR_Know represents the knowledge component of MAIR, and MAIR_Reg represents the regulation component.

Results from correlational analyses and multiple regression models for impulsive and strategic prediction bias can be seen in Tables 4.8 and 4.9. The regression models were not significant predictors of impulsive prediction bias, regardless of whether they included performance, $R^2 = .046$, $F_{(3, 54)} = 0.87$, $p = .460$, or not, $R^2 = .022$, $F_{(2, 55)} = 0.61$, $p = .546$. MAIR knowledge (Fig. 4.5A), regulation (Fig. 4.5C), and performance (Fig. 4.5E) all failed to predict impulsive prediction bias significantly. Similarly, no regression model was significant in predicting strategic prediction bias, regardless of whether they included performance, $R^2 = .069$, $F_{(3, 59)} = 1.45$, $p = .237$, or not, $R^2 = .042$, $F_{(2, 60)} = 1.31$, $p = .278$. MAIR knowledge did not show any association with bias (Fig. 4.6A), whereas all associations of MAIR regulation (Fig. 4.6C) and performance (Fig. 4.6E) with bias were not significant. Overall, neither MAIR components, nor running performance were significant predictors of impulsive and strategic running bias, mirroring correlational analyses.

Figure 4.5

Scatterplots illustrating the relationships of MAIR components and performance with calibration bias and absolute accuracy for impulsive predictions.



Note. Panels A, C, and E show the associations of MAIR Knowledge, MAIR Regulation, and performance with bias. Panels B, D, and F show the associations of MAIR Knowledge, MAIR Regulation, and performance with absolute accuracy.

Table 4.9

Multiple regression coefficients for MAIR component and performance predictors on the outcome variables bias.

Prediction	**Performance inclusion**	**Coefficient**	***B***	***Beta***	***Std. Error***	***t***	***p***
Impulsive	No Performance	Intercept	32.51	—	23.87	1.36	.179
		MAIR_Know	-0.30	-0.11	0.51	-0.59	.557
		MAIR_Reg	-0.07	-0.05	0.26	-0.27	.790
	Performance included	Intercept	61.04	—	33.94	1.80	.078
		MAIR_Know	-0.54	-0.20	0.54	-0.99	.327
		MAIR_Reg	0.00	0.00	0.26	0.00	.997
		Performance	-0.08	-0.17	0.07	-1.18	.244
Strategic	No Performance	Intercept	13.69	—	14.47	0.95	.348
		MAIR_Know	0.27	0.15	0.32	0.84	.402
		MAIR_Reg	-0.24	-0.28	0.15	-1.57	.121
	Performance included	Intercept	32.65	—	20.42	1.60	.115
		MAIR_Know	0.11	0.06	0.34	0.33	.742
		MAIR_Reg	-0.19	-0.22	0.16	-1.22	.228
		Performance	-0.06	-0.18	0.04	-1.31	.196

Note. Predictors were entered at the same time in each multiple regression model. B and Beta represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. MAIR_Know is Metacognitive knowledge in running, and MAIR_Reg is Metacognitive regulation in running.

Absolute Accuracy. Results from correlational analyses and multiple regression models for impulsive and strategic prediction absolute accuracy can be seen in Tables 4.8 and 4.10. For impulsive predictions, no regression model was a significant predictor of absolute accuracy, regardless of whether they included performance, R^2 = .051, $F_{(3, 54)}$ = 0.96, p = .417, or not, R^2 = .041, $F_{(2, 55)}$ = 1.17, p = .318. Prior to accounting for performance variance, participants with higher MAIR knowledge scores exhibited a non-significant tendency to be more precise in their impulsive predictions than participants with lower scores (Fig. 4.5B). However, this tendency diminished after accounting for performance. MAIR regulation (Fig. 4.5D) and performance (Fig. 4.5F) were also not significant predictors of impulsive absolute accuracy.

Similarly, regression models failed to predict absolute accuracy for strategic predictions, regardless of whether they included performance, R^2 = .036, $F_{(3, 59)}$ = 0.74, p = .530, or not, R^2 = .006, $F_{(2, 60)}$ = 0.18, p = .832. Neither MAIR component contributed to strategic prediction precision (Fig. 4.6B & 4.6D). Slower performers exhibited a tendency to be less precise than faster performers (Fig. 4.6F), but this tendency did not reach significance. Overall, neither

MAIR component was successful in predicting running absolute accuracy across prediction types, mirroring findings from bias and correlational analyses.

Table 4.10

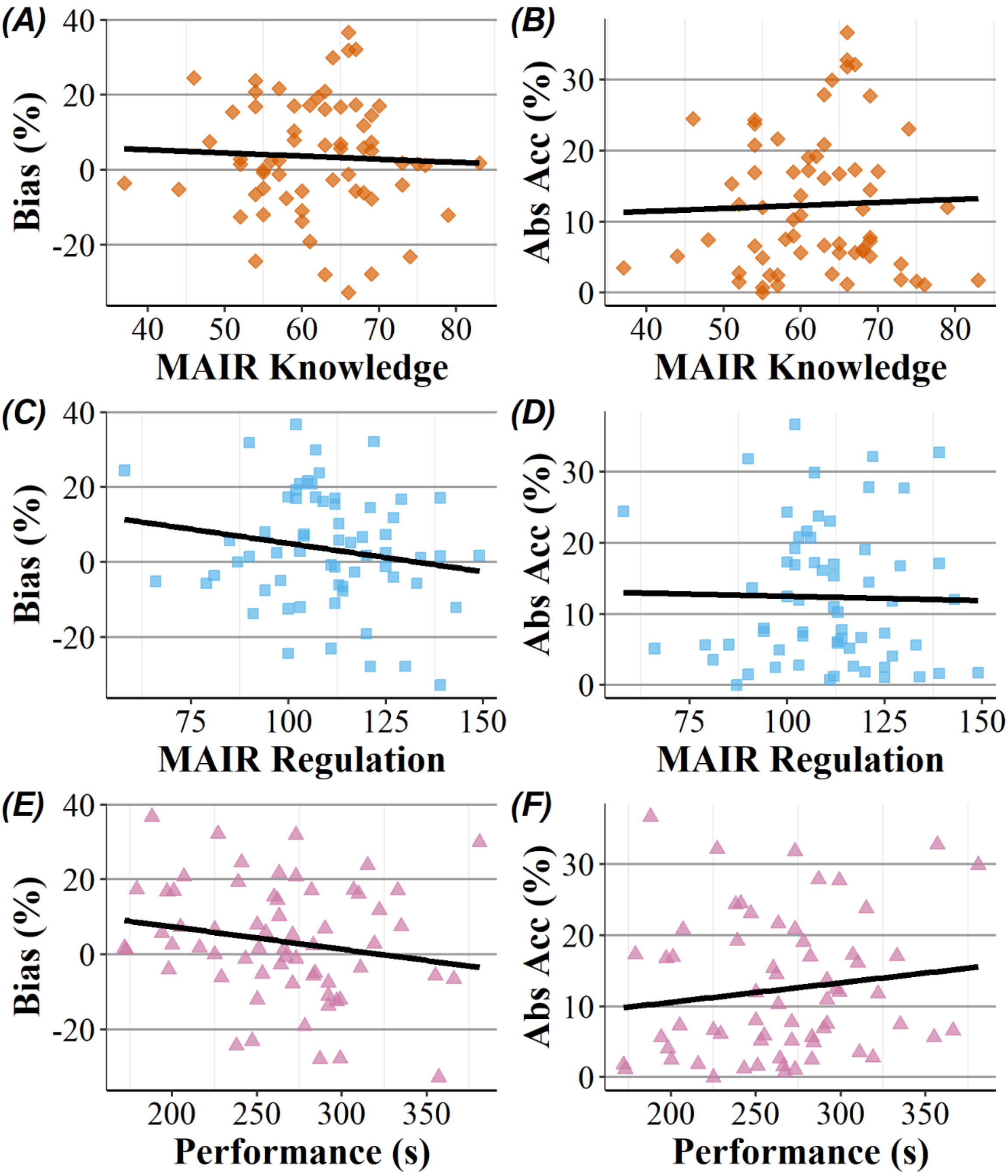
Multiple regression coefficients for MAIR component and performance predictors on the outcome variables absolute accuracy.

Prediction	Performance inclusion	Coefficient	*B*	*Beta*	*Std. Error*	*t*	*p*
Impulsive	No Performance	Intercept	25.51	—	17.81	1.43	.158
		MAIR_Know	-0.57	-0.27	0.38	-1.50	.139
		MAIR_Reg	0.24	0.23	0.19	1.25	.217
	Performance included	Intercept	11.75	—	25.51	0.46	.647
		MAIR_Know	-0.45	-0.22	0.41	-1.11	.273
		MAIR_Reg	0.21	0.20	0.20	1.05	.300
		Performance	0.04	0.11	0.05	0.76	.453
Strategic	No Performance	Intercept	10.84	—	9.28	1.17	.247
		MAIR_Know	0.12	0.11	0.20	0.58	.563
		MAIR_Reg	-0.05	-0.10	0.10	-0.53	.597
	Performance included	Intercept	-1.82	—	13.08	-0.14	.890
		MAIR_Know	0.22	0.20	0.22	1.03	.307
		MAIR_Reg	-0.09	-0.16	0.10	-0.85	.398
		Performance	0.04	0.19	0.03	1.36	.178

Note. Predictors were entered at the same time in each multiple regression model. B and Beta represent the unstandardized and standardized estimates of the coefficients respectively and Std. Error represents standard error of the mean of this estimate. *t* and *p* represent the test statistic and *p*-value associated with the corresponding predictors. MAIR_Know is Metacognitive knowledge in running, and MAIR_Reg is Metacognitive regulation in running.

Figure 4.6

Scatterplots illustrating the relationships of MAIR components and performance with calibration bias and absolute accuracy for strategic predictions.



Note. Panels A, C, and E show the associations of MAIR Knowledge, MAIR Regulation, and performance with bias. Panels B, D, and F show the associations of MAIR Knowledge, MAIR Regulation, and performance with absolute accuracy.

4.3.4 Discussion of Study 5

The aim of Study 5 was to examine whether self-reports of running-specific metacognition would predict calibration in a familiar running task.

Contrary to my expectations of running-specific self-reports predicting running calibration, there was no relationship between self-reports of running metacognition and running calibration. Neither MAIR knowledge, nor MAIR regulation were significant predictors of running bias and absolute accuracy. This result was not consistent with Nietfeld's (2003) finding of running monitoring accuracy being associated with higher self-reports of running metacognition. Furthermore, it did not support the suggestion that metacognition questionnaires that are specific to the task examined should be more likely to correlate with online measures of metacognition (Schellings, 2011; Schellings et al., 2013). Findings from Study 5 mirrored those of Study 4, suggesting that we cannot use self-reports of exercise metacognition to predict calibration reliably, regardless of task familiarity and questionnaire specificity. Consequently, there appears to be a dissociation between offline and online measures of metacognition, meaning that we cannot use one to inform the other in a reliable manner. This raises the question of whether online measures of metacognition collected in non-exercise domains might be more effective in informing us about exercise calibration instead.

4.4 STUDY 6 – COGNITIVE AND HIFME CALIBRATION

4.4.1 Study specifics

The absence of a significant relationship between metacognition self-reports and exercise calibration observed in Studies 4 and 5 indicate a dissociation between offline and online measures of metacognition, but cannot inform us on whether there are dissociations between online measures of metacognition in different domains. Findings from Section 4.2.3.1 indicated that self-reports of cognitive metacognition can provide us with information about self-reports of exercise metacognition, and vice versa. Furthermore, as discussed in Section 4.1, results from Jonker and colleagues (2010, 2011) also support the transfer of metacognitive processes across exercise and cognitive domains. This raises the question of whether online measures of metacognition can exhibit a similar capacity for transfer across domains.

There has also been empirical research to suggest that calibration generalises across different cognitive domains (e.g. Carpenter et al., 2019; Mazancieux et al., 2020; Rouault, McWilliams, Allen, & Fleming, 2018). Similarly, Arbuzova and colleagues (2020) found that metacognitive efficiency correlated across visual and motor conditions of a ball-throwing task, indicating domain-generality of processes involved in monitoring different types of sensory feedback (see Section 4.1). Based on these findings, it is possible that metacognitive processes contributing to performance judgments are implemented consistently across cognitive and exercise domains, leading to calibration correlations between them. In Study 6, I tested this possibility by examining the presence of a relationship between memory recognition and HIFME calibration. Based on the findings discussed above, I expected that memory recognition bias and absolute accuracy would correlate with HIFME bias and absolute accuracy respectively.

4.4.2 Methods

4.4.2.1 Participants

Participants in the present analysis came from the sample used in Study 3, which I presented in Chapter 3 (see Section 3.2.2.1). Along with non-finishers and outliers, I also excluded one participant from the present study, whom I had not excluded from the analysis in Study 3, as they did not complete the memory recognition task due to equipment malfunction. Overall, I analysed data from a sample of fifty-two participants (24 men and 28 women; M_{age} = 24.6

years old, *SD* = 5.0 years old) for the AMRAP workout, and data from a sample of fifty-five participants (27 women and 28 men; M_{age} = 24.7 years old, *SD* = 5.1 years old) for the Rounds workout.

This study received approval by the St Andrews School of Psychology & Neuroscience Ethics Committee (Ethics approval code: PS13328; see Appendix 8.1.3). Participants were compensated at a rate of £5/hour.

4.4.2.2 Memory recognition task

I generated a list of 1962 random nouns of fairly similar frequency from the English Lexicon Project database (Balota et al., 2007). For each participant session, I randomly selected 50 nouns from the list. Participants were presented with one word at a time, which stayed on the screen for 1.5 seconds. Participants then had to provide their Judgments of Learning (JOLs) for each noun, rating the likelihood they would be able to recognise it in the future on a scale from 0 to 100. I divided JOL scores in increments of 20 (i.e. 0%, 20%, 40%, 60%, 80%, and 100%). As soon as all 50 words had been presented, participants had the opportunity to rest, and were then presented with a list of 100 words. This list contained the previously presented 50 words, as well as another new 50 randomly generated words from the dictionary list. Participants then had to decide whether each word was in the previous list, pressing right arrow for 'old' and left arrow for 'new'. After making a judgment for each word, participants were asked to report their confidence on their answer being correct on a scale from 1 (low confidence) to 3 (high confidence). Prior to completing the test trials, participants also completed a practice run of the task, with the initial list containing five words, and the second list containing ten.

4.4.2.3 HIFME workouts

Details for the AMRAP and Rounds workouts are identical to those used in Study 3, and can be found in Section 3.2.2.2.

4.4.2.4 Design

The aim of the present analysis was to examine the relationship between HIFME and cognitive calibration. I thus explored the associations of AMRAP and Rounds bias and absolute accuracy with JOL and confidence rating bias and absolute accuracy in the memory recognition task. I calculated HIFME bias and absolute accuracy percentages as described in

Section 3.2.2.4. I calculated calibration measures in a different way for the memory recognition task, as it contained multiple trials. For JOL calibration, I divided JOL scores by 100 to match the scale of recognition accuracy scores (0 for incorrect and 1 for correct). I calculated bias and absolute accuracy percentages for JOLs as seen in (1) and (2). In the calibration calculations for confidence judgments, I assigned confidence ratings of 3 the value of 1 (i.e. high confidence), ratings of 2 the value of 0.5 (i.e. moderate confidence), and ratings of 1 the value of 0 (i.e. low/no confidence). I calculated bias and absolute accuracy percentages for confidence ratings as seen in (3) and (4).

$$BiasJOL = [\frac{1}{50}\sum_{i=1}^{50}\big(JOL(i) - Accuracy\ score(i)\big)] \text{ x } 100 \tag{1}$$

$$AbsAccJOL = [\frac{1}{50}\sum_{i=1}^{50}|JOL(i) - Accuracy\ score(i)|]\ x\ 100 \tag{2}$$

$$BiasConf = [\frac{1}{100}\sum_{i=1}^{100}(Confidence(i) - Accuracy\ score(i))]\ x\ 100 \tag{3}$$

$$AbsAccConf = [\frac{1}{100}\sum_{i=1}^{100}|Confidence(i) - Accuracy\ score(i)|]\ x\ 100 \tag{4}$$

4.4.2.5 Procedure

Details on the procedure implemented in this study can be found in Section 3.2.2.5. An addition to the procedure described there is that participants completed the memory recognition task after they had provided me with their informed consent, and before I asked them to complete the exercise experience questionnaire (and the HIFME workouts).

4.4.3 Results

I conducted correlations to examine associations between HIFME workouts and the memory recognition task for every outcome variable (i.e. performance, metacognitive judgments, bias, and absolute accuracy). Correlations between the AMRAP and the HIFME workouts for performance predictions, performance, bias, and absolute accuracy can be seen in Section 3.2.3.3.3.

4.4.3.1 Memory recognition & HIFME performance, predictions, and calibration

For the memory recognition task, I present descriptive statistics for performance, metacognitive judgments, and calibration based on the sample I used for comparisons with the Rounds workout. On average, participants were correct in 87.75% (*SD* = 7.82%) of the trials for which they made JOLs, and their mean JOL scores were 62.47% (SD = 16.43%). Across trials (i.e. trials for both new and old items), participants were correct in 86.76% (SD = 6.92%) of cases, and their mean post-trial confidence ratings were 2.49 (SD = 0.32). In the memory recognition task, positive JOL and confidence bias values indicate overconfidence, whereas negative values indicate underconfidence. Participants were thus significantly underconfident in both their JOLs (M = -25.64%, SD = 18.51%), $t_{(54)} = -10.27$, $p < .001$, $d = -1.39$, and their confidence ratings (M = -12.41%, SD = 14.88%), $t_{(54)} = -6.19$, $p < .001$, $d = -0.83$, as they performed better than they anticipated. Mean absolute accuracy scores were 40.62% (SD = 12.79%) for JOLs, and 25.72% (SD = 11.54%) for confidence ratings.

Participants in the sample used for AMRAP analyses in Study 6 completed an average of 195 reps (SD = 46 reps), and predicted an average of 186 reps (SD = 59 reps). Participants exhibited a non-significant tendency towards underconfidence in their AMRAP predictions (M = -3.75%, SD = 23.92%; negative values indicate underconfidence in the AMRAP workout), $t_{(51)} = -1.13$, $p = .264$, $d = -0.16$. The mean absolute accuracy of AMRAP predictions was 18.82% (SD = 15.01%). Participants in the sample used for Rounds analyses in Study 6 completed the workout in an average of 515 s (SD = 123 s), and their mean predicted finish times were 536 s (SD = 151 s). Participants exhibited a non-significant tendency to be underconfident in their Rounds predictions (M = 5.05%, SD = 22.85%; positive values indicate underconfidence in the Rounds workout), $t_{(54)} = 1.64$, $p = .107$, $d = 0.22$. The mean absolute accuracy of Rounds predictions was 17.85% (SD = 14.95%).

4.4.3.2 Performance comparisons across tasks

For performance comparisons, I used overall performance for the memory task, and not just the "old" items for which participants provided their JOLs. Neither AMRAP, nor Rounds performance, correlated with memory recognition performance—AMRAP: $r_{(50)} = .06$, $p = .693$; Rounds: $r_{(53)} = -.18$, $p = .195$. Therefore, there was no association between HIFME and memory recognition performance.

4.4.3.3 Predictions & Confidence comparisons across tasks

There was a moderate positive correlation between JOLs and confidence ratings for the memory recognition task, $r_{(53)} = .48$, $p < .001$. Participants who predicted they would be more likely to recognise the words presented in the future were also more confident in the accuracy of their answers. Predictions for the AMRAP workout did not correlate with either JOLs, $r_{(50)}$

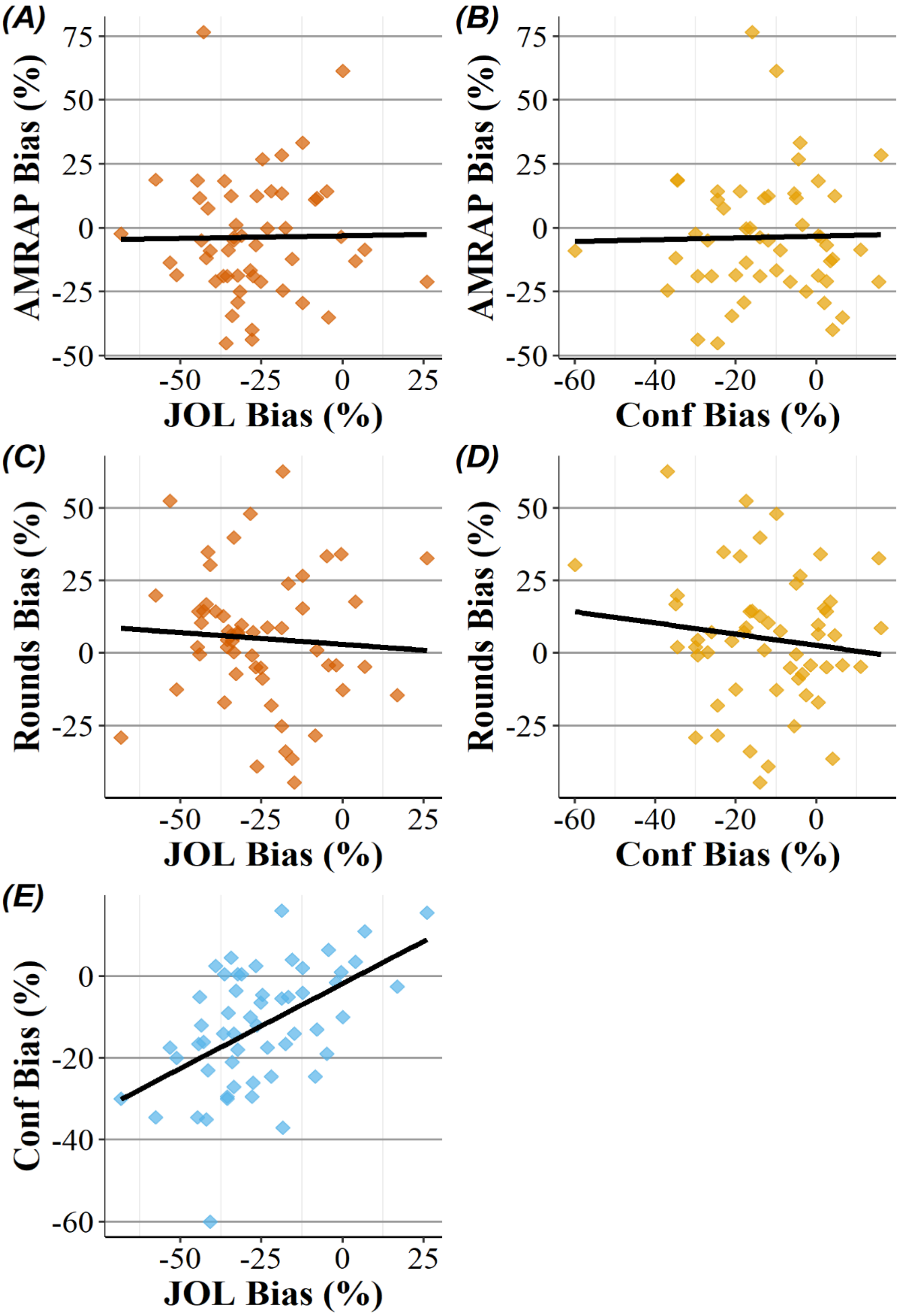
= .00, p = .999, or confidence ratings, $r_{(50)}$ = .07, p = .612. Similarly, predictions for the Rounds workout did not correlate with JOLs, $r_{(53)}$ = -.08, p = .586, but they did exhibit a non-significant correlation with confidence judgments, $r_{(53)}$ = -.25, p = .070. Participants who predicted faster finish times for the Rounds workout had a minor tendency to also be more confident in the accuracy of their memory recognition responses. Overall, these findings suggest that there is no reliable relationship between HIFME predictions and memory recognition JOLs and confidence ratings.

4.4.3.4 Calibration comparisons across tasks

Bias. There were no significant correlations between cognitive and HIFME bias. JOL bias did not correlate with HIFME bias in either the AMRAP, $r_{(50)}$ = .02, p = .917 (Fig. 4.7A), or the Rounds, $r_{(53)}$ = -.07, p = .634 (Fig. 4.7C), workouts. Similarly, confidence rating bias did not correlate with HIFME bias in either workout—AMRAP: $r_{(50)}$ = .02, p = .877 (Fig. 4.7B); Rounds: $r_{(53)}$ = -.13, p = .360 (Fig. 4.7D). In contrast, there was a moderate positive correlation for bias between JOLs and confidence ratings, $r_{(53)}$ = .52, p < .001, indicating that participants exhibited consistent patterns of bias across metacognitive judgments for the memory recognition task (Fig. 4.7E). Therefore, though participants showed a moderate tendency for their bias to be consistent across prospective and retrospective metacognitive judgments for the memory recognition task, cognitive bias was not associated with HIFME bias, indicating domain dissociation.

Figure 4.7

Scatterplots illustrating the relationships between memory recognition and HIFME bias.

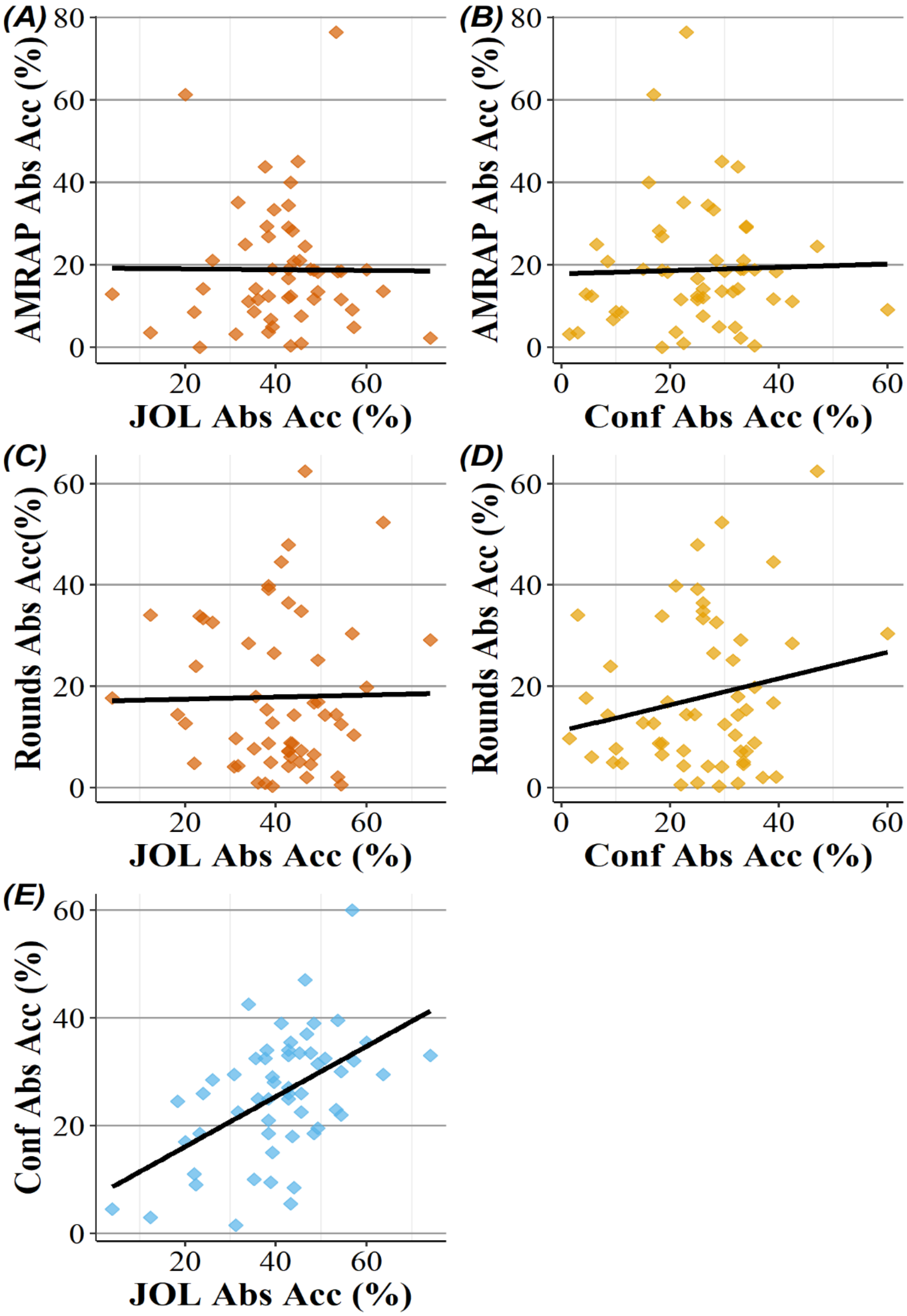


Note. Panels A and B show the associations of JOL and confidence rating bias with bias in the AMRAP workout. Panels C and D show the associations of JOL and confidence rating bias with bias in the Rounds workout. Panel E illustrates the relationship between JOL and confidence rating bias.

Absolute Accuracy. JOL absolute accuracy did not correlate with either AMRAP, $r_{(50)} = -.01$, $p = .957$ (Fig. 4.8A), or Rounds, $r_{(53)} = .02$, $p = .900$ (Fig. 4.8C), absolute accuracy. Similarly, confidence rating absolute accuracy did not correlate with absolute accuracy for either workout—AMRAP: $r_{(50)} = .03$, $p = .829$ (Fig. 4.8B); Rounds: $r_{(53)} = .20$, $p = .143$ (Fig. 4.8D). Conversely, participants exhibited a moderate positive correlation between their JOL and confidence rating absolute accuracy, $r_{(53)} = .52$, $p < .001$, indicating that participants who made precise JOLs were also more likely to make precise confidence ratings (Fig. 4.8E). These results mirror those for bias, as participants were moderately consistent in their precision across metacognitive judgments for the memory recognition task, but there were no correlations between memory recognition and HIFME absolute accuracy.

Figure 4.8

Scatterplots illustrating the relationship between memory recognition and HIFME absolute accuracy.



Note. Panels A and B show the associations of JOL and confidence rating absolute accuracy with absolute accuracy in the AMRAP workout. Panels C and D show the associations of JOL and confidence rating absolute accuracy with absolute accuracy in the Rounds workout. Panel E illustrates the relationship between JOL and confidence rating absolute accuracy.

4.4.4 Discussion of Study 6

The aim of Study 6 was to examine the extent to which bias and absolute accuracy in a memory recognition task correlated with bias and absolute accuracy in two HIFME workouts.

Contrary to my prediction of cognitive calibration exhibiting an association with HIFME calibration, bias and absolute accuracy in cognition and exercise did not exhibit any correlations with each other. Though participants showed a moderate tendency to be consistent in their bias and precision across prospective (i.e. JOLs) and retrospective (i.e. confidence ratings) metacognitive judgments in the memory recognition task, this consistency did not extend to HIFME calibration. These results were not in line with those of Arbuzova and colleagues (2020), suggesting that exercise calibration is less likely to correlate with cognitive calibration than motor calibration is. Nonetheless, it should be noted that Arbuzova and colleagues examined calibration using metacognitive efficiency, whereas I used bias and absolute accuracy. Additionally, participants in their study produced local metacognitive judgments (i.e. they gave confidence ratings for each trial) across conditions, whereas participants in Study 6 produced local metacognitive judgments for the memory recognition task and global estimates of performance for the exercise tasks (i.e. one metacognitive judgment per task). Differences in methodology could have thus contributed to the discrepancies between studies. Interestingly, I also observed no relationships between performance and metacognitive estimates across the cognitive and exercise domains, indicating a general dissociation between the two domains that was not exclusive to calibration. It is worth investigating the extent to which different cognitive tasks are more closely related to exercise tasks, and whether this might affect calibration comparisons across domains.

4.5 GENERAL DISCUSSION

Chapter 4 examined the extent to which we can use offline measures of metacognition and cognitive calibration to predict exercise calibration. To achieve this, I analysed and presented data from three studies. In Study 4, I explored whether general exercise self-reports of metacognition would be associated with calibration for a novel HIFME workout. Given the equivocality of previous cognitive findings that used non-task-specific questionnaires, I did not make a specific prediction for Study 4 results. In Study 5, I investigated the relationship between running-specific self-reports of metacognition and running calibration, recruiting participants with previous running experience. Because I used a task-specific questionnaire, I expected that running self-reports of metacognition would predict running calibration. In Study 6, I focused on online measures of metacognition, and tested the extent to which calibration in a memory recognition task can predict calibration in the two HIFME workouts previously presented in Study 3 in Chapter 3. I predicted that cognitive calibration would exhibit a positive correlation with HIFME calibration.

Self-reports of general exercise metacognition failed to predict HIFME calibration in Study 4. MAIE knowledge and regulation scores did not exhibit a relationship with either HIFME bias or absolute accuracy. Furthermore, though higher MAIE regulation predicted better HIFME performance, neither MAIE component contributed to performance predictions. Overall, results illustrated that self-reports of exercise metacognition were not associated with either HIFME predictions, or calibration. They were thus inconsistent with research that has observed a positive correlation between domain-general self-reports of academic metacognition and cognitive calibration (Jang et al., 2020; Schraw, 1997; Tobias et al., 1999), but in line with studies that have not observed such relationships (Gutierrez & Schraw, 2015; Jacobse & Harskamp, 2012; Saraç & Karakelle, 2012; Schraw & Dennison, 1994; Sperling et al., 2004; Zepeda et al., 2015). Nonetheless, the absence of significant findings in Study 4 does not necessarily show that we cannot use metacognition self-reports to predict exercise calibration, but rather that self-reports might have to be specific to the calibration task used, and that participants should be familiar with the exercise modality implemented. This led to Study 5, where participants had previous experience with the exercise modality tested, and the metacognitive questionnaire used was specific to it.

Contrary to my expectations, self-reports of running metacognition did not predict running calibration for either impulsive, or strategic predictions. Similar to Study 4, MAIR knowledge and regulation were not associated with bias or absolute accuracy for either prediction type. Interestingly, unlike Study 4, where, after accounting for variance shared by the two MAIE components, only metacognitive regulation predicted HIFME performance, in Study 5, it was metacognitive knowledge that predicted performance. Furthermore, participants with higher running metacognitive knowledge exhibited significant and non-significant tendencies to make faster impulsive and strategic predictions respectively, whereas neither metacognitive component contributed to performance predictions in Study 4. This suggests that metacognitive knowledge regarding a physical activity can contribute to predicting performance when there is a priori experience with the activity in question.

Nonetheless, Study 5 findings did not support the suggestion that task-specific metacognition self-reports are more likely to correlate with online measure of metacognition (Schellings, 2011; Schellings et al., 2013). They were also incongruent with Nietfeld's (2003) findings of higher monitoring accuracy in runners who reported higher running metacognition. We could attribute this discrepancy between results to Nietfeld not examining calibration directly. However, it is also possible that the MAIR used here was still too broad to show a clear connection with running calibration—especially since Nietfeld narrowed down his questionnaire to only a few and specific items that targeted running monitoring. This renders the degree of questionnaire specificity required for a metacognition inventory to predict calibration unclear. It is possible that simply adjusting a questionnaire, such as MAI, to a modality by making its items specific to it is not sufficient to produce a reliable relationship between offline and online measures of metacognition (Jacobse & Harskamp, 2012). In that case, it might be preferential to select only a limited number of items that are specific to the metacognitive processes involved in the online measures of metacognition examined. If this suggestion is empirically supported, then developing and using short and specific metacognition self-reports will allow us to predict exercise calibration.

Results in Study 6 were also not consistent with my hypothesis, as JOL and confidence rating bias and absolute accuracy in the memory recognition task did not correlate with bias and absolute accuracy in either HIFME workout. These results were not in line with Arbuzova and colleagues' (2020) findings of metacognitive efficiency correlations between motor and cognitive task conditions. This discrepancy could stem from differences in the extent to which bias and absolute accuracy show similar domain generality as other calibration measures, such as metacognitive efficiency. However, such metacognitive measures need multiple trials for their calculation (see Section 1.2.3.2 for review), so it is not possible to calculate them in exercise studies that collect data from single trials. Furthermore, the exercise tasks used in Study 6 were arguably more complex than the motor conditions in the ball-throwing task used by Arbuzova and colleagues (2020), which could render it more difficult for calibration patterns to transfer across domains. In fact, bias and absolute accuracy did not correlate between the two HIFME workouts that belonged to the same domain (see Section 3.2.3.3.3), so it is less likely that they would correlate with calibration measures in a task in a completely different domain. Interestingly, there was also no significant relationship between the cognitive and exercise modalities for performance and metacognitive judgments (i.e. performance predictions, JOLs, and confidence ratings), illustrating a general dissociation between the memory recognition and HIFME tasks.

Another factor to consider is that participants in the study conducted by Arbuzova and colleagues (2020) gave local metacognitive judgments across the motor and cognitive conditions, as they provided confidence ratings for each trial they completed. In contrast, participants in Study 6 gave local metacognitive judgments only for the memory recognition task. Since they produced one performance estimate for each workout, we can interpret their metacognitive judgments for the exercise tasks as being global estimates of performance instead. Recent research has shown that, though there are similarities between global and local metacognitive judgments, there are also differences (Händel, de Bruin, & Dresel, 2020; Karst, Dotzel, & Dickhäuser, 2018; Rouault, Dayan, & Fleming, 2019; Rouault & Fleming, 2020), suggesting that consistency in the use of metacognitive judgment types is warranted when examining relationships in metacognition between different tasks. It is thus possible that by using global estimates of performance across the exercise and cognitive tasks, I could have increased the likelihood of Study 6 producing results in line with those of Arbuzova and colleagues (2020). For all the above reasons, we should further examine the relationship between cognitive and exercise calibration by using less complex exercise tasks, measuring cognitive calibration using tasks that are more closely related to the exercise activities implemented, and by asking participants to provide global estimates of performance across tasks.

In Study 4, I also collected data on academic metacognition self-reports to examine whether MAIE scores would correlate with MAI scores. Contrasting calibration findings from Study 6 and in accordance with my predictions and previous research on the domain-generality of metacognition reports across exercise and academic domains (Jonker et al., 2010, 2011; Mccardle, 2015), the MAIE and MAI questionnaires exhibited moderate-to-strong positive correlations with each other. Higher scores of metacognitive knowledge and regulation in exercise were associated with higher scores of metacognitive knowledge and regulation in learning respectively. Though all components correlated with each other, the correlations were stronger for congruent components across questionnaires (e.g. MAIE knowledge correlated more strongly with MAI knowledge than with MAI regulation). These results suggest that individuals who report that they engage in metacognitive behaviour in exercise are more likely to also report similar metacognitive engagement in academia, and vice versa. Nonetheless, it should be noted that only participants with previous exercise experience participated in Study 4, so this relationship might be weaker among students who do not exercise. Furthermore, I developed MAIE based on MAI, so the similarity between the two questionnaires likely contributed to the correlations observed. Though this methodology was similar to the one used by Nietfeld (2003), it would be interesting to examine the reliability of the present finding using a metacognitive questionnaire for exercise that has been developed independently of MAI. Finally, though components from the two questionnaires correlated significantly with each other, the correlations were only moderate-to-strong, suggesting that domain-specific factors also make a contribution to metacognition self-reports.

The studies presented in the present chapter have important implications for assessing exercise calibration. Self-reports of metacognition and cognitive calibration failed to predict exercise calibration, regardless of whether the questionnaires were specific to the exercise modality used or not. Furthermore, previous task familiarity did not affect these results. It thus appears that we cannot use offline measures of metacognition to make reliable assessments regarding exercise calibration. This could be due to numerous reasons. Participants may tend to exaggerate their use of metacognition in questionnaires (Tobias et al., 1999), leading to discrepancies between their perceptions of metacognitive engagement and how much they actually implement metacognitive knowledge and skills when predicting performance. Furthermore, better and more experienced athletes might have highly automatized metacognitive processes, which they do not take in consideration when completing metacognitive questionnaires (Harrison & Vallin, 2018). It should also be noted that using broad questionnaires with multiple items, such as MAI, might not allow questionnaires to be specific enough to the metacognitive processes that contribute to calibration. Since Nietelfd (2003) observed a relationship between running monitoring and relevant metacognition self-reports using only 10 items, it is interesting to test whether devising questionnaires that only use a small number of items that are specific to exercise calibration will lead to different result to the present studies.

Results from Study 6 suggest that cognitive calibration is not effective in informing us about exercise calibration, though, as discussed above, further research using consistent metacognitive judgments, as well as different cognitive and exercise tasks, is required to evaluate the extent to which the present results are reliable. Consistent replication of the present findings with different methodologies would indicate a reliable dissociation in the implementation of metacognitive processes between the domains of exercise and cognition when assessing performance. In contrast, Study 4 showed that we could use self-reports of academic metacognition to make inferences about self-reports of exercise metacognition, and vice versa, in university students who exercise consistently. Nonetheless, there appear to also be domain-specific factors that contribute to metacognition self-reports, whilst this relationship might be weaker for students who do not engage in sports, or athletes who do not engage in cognitive/academic work.

The research presented in the chapter was not without limitations. I only developed MAIE and MAIR for the present studies and I did not test their reliability and validity extensively. Though the examination of reliability and validity for exercise metacognition questionnaires was outside the scope of the present studies, it would be interesting to know whether the questionnaires were measuring what they were supposed to measure, and whether they would be able to produce consistent scores across time. This would in turn provide us with information regarding the validity and reliability of the results from the present studies. Another potential limitation is that I collected calibration data in Study 5 using specific instructions for impulsive and strategic predictions (I analyse and present the effect of this manipulation in Study 8 in Section 5.3), rather than simply asking participants to provide their predictions using any method they wanted. However, impulsive and strategic predictions are common ways in which athletes predict performance, so it is unlikely that they altered results considerably.

4.6 CONCLUSION

Chapter 4 examined the extent to which self-reports of exercise metacognition and cognitive calibration can predict exercise calibration. The results suggested that offline measures and calibration patterns from other domains are not effective in informing us about athlete calibration. This means that using other metacognition measures to assess calibration in an exercise task is likely ineffective. This was surprising, as, given that calibration is a measure of metacognitive monitoring and that numerous metacognitive processes are thought to be domain-general, we would expect different measures of metacognition to exhibit relationships with each other. Overall, though it is worth investigating whether we would observe the same results using different exercise and cognitive tasks, and making self-report questionnaires very specific to calibration processes and the task examined, we can deduce that assessing calibration using other metacognitive measures is a less straightforward task than we would expect. Nonetheless, results in the present chapter do not suggest that metacognition does not play a role in role in understanding calibration altogether. Instead, they highlight the need to examine the extent to which metacognition can contribute to calibration in different ways. This is the aim of Chapter 5, where I investigate whether we can utilise metacognitive manipulations to improve exercise calibration.

CHAPTER 5: PREDICTION GUIDANCE EFFECTS ON HIFME AND RUNNING CALIBRATION

5.1 INTRODUCTION & RATIONALE

In Chapter 4, I focused on the link between calibration and trait metacognition. Since calibration is a measure of metacognitive monitoring accuracy (see Section 1.2.3.2), I examined whether other measures of metacognition (i.e. self-reports and calibration data collected in a different modality) could inform us about exercise calibration. Results from the first two studies illustrated that metacognition self-reports cannot provide us with reliable information about exercise calibration. In the same vein, calibration in a memory recognition task was not associated with calibration in a HIFME task. Overall, though these results suggest that we might not be able to use metacognition self-reports and cognitive calibration to predict exercise calibration, they do not constitute evidence for the presence of a dissociation between metacognition and calibration. Differences between the ways in which we collect metacognitive data using offline and online measures, as well as differences between cognitive and exercise tasks, can explain the absence of significant findings instead. Calibration should still have a close relationship with metacognition, even if this is not visible in comparisons between self-report scores and calibration.

Accordingly, though cognitive research has been inconsistent in finding a relationship between online and offline measures of metacognition (see Section 1.3.2.1 for review), it has been consistent in finding evidence of metacognitive training/interventions improving calibration (see Section 1.3.2.2 for review). This discrepancy in findings bolsters the suggestion that the close relationship between metacognition and calibration is not apparent in studies using self-reports because of issues pertaining to the methodologies used, e.g. individuals reporting high metacognitive knowledge and regulation, which do not reflect actual behaviour, leading to poorer-than-expected calibration (Tobias et al., 1999). It thus appears that to better understand the relationship between metacognition and calibration in exercise and how we can use the former to improve the latter, we need to conduct more research that manipulates metacognitive behaviour directly (e.g. by training participants to better monitor their behaviour) and examines the manipulation's effects on exercise calibration. This was the aim of Chapter 5, where I tested the extent to which instructing participants to engage in strategic thinking when making their predictions would lead to better calibration than instructing them to make impulsive, non-strategic, predictions.

Research on the effects of metacognitive manipulations on exercise calibration has been scarce (see Section 1.3.2.2). Exercise studies have only examined the effects of self-regulation training (which also targeted metacognitive processes) on calibration in physical education settings. The Kolovelonis lab (Kolovelonis, Goudas, & Dermitzaki, 2012; Kolovelonis et al., 2013) found that self-regulation training did not lead to improved basketball dribbling calibration relative to baseline measures, mirroring results from participants in the control group who did not receive self-regulation training. The researchers suggested that the absence of significant findings resulted from the interventions in the experimental groups not focusing on metacognitive processes that contribute to calibration sufficiently. To address this limitation, Kolovelonis and colleagues (2020) examined the effects of self-regulation training that specifically targeted metacognitive processes closely associated with calibration (i.e. setting own practice goals, self-recording goals and performance, engaging in self-talk, self-reflecting and self-evaluating own performance, and making performance attributions) on basketball shooting calibration. Contrasting previous findings, participants who received metacognitive training exhibited higher prediction precision compared to their baseline measurements. No such improvements were present in the control group that did not receive self-regulation training. Therefore, there is preliminary evidence to suggest that metacognitive training is an effective tool for the facilitation of exercise calibration. However, these studies also illustrate the importance of devising effective manipulations that target calibration-related metacognitive processes. Since research on the effects of metacognitive interventions on sports calibration is limited, it becomes essential to explore which types of interventions are most effective in improving exercise calibration.

There are numerous types of metacognitive manipulations we can use to facilitate calibration. One way is to provide participants with training that aims to improve their metacognitive abilities and capacity to assess their performance, e.g. though extensive instructions on when and how to best monitor, review, and evaluate their performance and progress (e.g. Gutierrez & Schraw, 2015). Another way is by simply instructing participants to engage in certain types of metacognitive behaviour when they practice a task (e.g. self-recording and self-evaluating performance), without providing them with more extensive metacognitive training (e.g. Kolovelonis et al., 2020). Researchers can also opt to provide participants with feedback on their metacognitive, rather than task, performance to train them in assessing the extent to which their metacognitive judgments reflect their performance accurately (e.g. Carpenter et al., 2019). The above methods are examples of how researchers have previously implemented metacognitive manipulations to optimise metacognitive behaviour and calibration, and which can thus inform us on how to best improve metacognitive function and calibration in athletes.

Nonetheless, even individuals who possess good metacognitive skills and extensive metacognitive knowledge can produce inaccurate performance estimates, if they do not engage in metacognitive behaviour when doing so. In the studies I presented in the previous chapters (excluding Study 5 where I provided participants with specific instructions on how to make their predictions), participants often described their reasoning for their predictions verbally during data collection. Some participants engaged in extensive metacognitive thinking to produce their predictions, whereas others indicated that their predictions were simply the "first number that came to mind" and that they had not thought about them. Though I did not collect quantitative data to assess this observation empirically, given previous cognitive and exercise research on the relationship between metacognitive behaviour and calibration (see Section 1.3.2.2), I would expect participants who made strategic predictions to be better calibrated than participants who were impulsive and less strategic in their estimates. This is because the former should have been more likely to take full advantage of their metacognitive skills and knowledge than the latter. The main aim of Chapter 5 was to test this prediction experimentally.

To examine the effect of engaging in metacognitive or impulsive behaviour when producing performance estimates on exercise calibration, I conducted two studies. In these two studies, I manipulated behaviour engagement when estimating prospective performance by providing participants with either strategic or impulsive prediction guidance. Prediction guidance referred to instructing participants on how to make their predictions, which constitutes a minimal and time-effective metacognitive intervention. If athletes can exhibit improved prediction calibration simply by receiving metacognitive instructions on how to make their predictions, then coaches, fitness instructors, and event organisers will benefit from using such instructions when they ask athletes to make performance estimates. Furthermore, results on this manipulation could better inform us regarding the types of metacognitive interventions that can facilitate exercise, and potentially cognitive, calibration.

For strategic predictions, participants made their performance estimates after receiving instructions prompting them to engage in metacognitive strategic thinking. For impulsive predictions, participants made their estimates after receiving instructions prompting them to be impulsive, and not strategic. I provided participants with impulsive instructions in the control conditions across studies to ensure that participants in these conditions did not engage in strategic thinking. Had I simply not provided them with any instructions, then it is likely that some participants would have engaged in strategic thinking, thereby reducing the study's statistical power to observe differences between the strategic and the non-strategic, control conditions. I anticipated that strategic instructions would lead to better calibration than impulsive instructions across studies.

5.2 STUDY 7 – INSTRUCTIONS & HIFME CALIBRATION

5.2.1 Study specifics

The aim of Study 7 was to examine whether providing inexperienced participants with specific instructions on how to make strategic performance predictions for a HIFME workout would lead to better calibration than instructing participants to make impulsive predictions. To test this, I randomly allocated participants in strategic and impulsive groups, with participants in the former receiving strategic instructions, and participants in the latter receiving impulsive instructions. Neither group had previous HIFME experience to ensure lack of familiarity with HIFME performance assessment processes, and thus avoid experience confounds in the results. Using an unfamiliar exercise modality also rendered potential results even more interesting, as they would indicate whether coaches and fitness instructors could use simple prediction instructions to facilitate performance awareness in athletes who have not participated in a specific athletic activity before.

For the impulsive group, I instructed participants to make their predictions based on their "gut feeling" and to avoid engaging in strategic thinking. For the strategic group, I asked participants to be "strategic" in their predictions by using a chunking-like strategy, i.e. I instructed them to think about the time required to complete a workout round to estimate total prospective performance (see Section 5.3.2.3 for the verbatim instructions). Chunking in sports refers to breaking down an exercise task in smaller components, and athletes have often reported its use to regulate fatigue and performance during endurance exercise (Brick, Campbell, et al., 2016; Brick, MacIntyre, et al., 2016; Brick et al., 2015). Because of its reported prevalence, I expected that participants would be familiar with the chunking strategy as a way of thinking about athletic performance, and that it would be successful in eliciting effective metacognitive thinking. Overall, I expected that participants in the strategic group would exhibit better calibration (i.e. lower bias and higher precision) than participants in the impulsive group.

5.2.2 Methods

5.2.2.1 Participants

Participants were recreational athletes who reported no previous HIFME experience, i.e. had either never engaged in HIFME or their cumulative HIFME experience was less than a month (meaning that even if they had tried HIFME before, they were still eligible to participate if they had not engaged in it consistently—e.g. for more than a total of ten hours). The criteria for HIFME and non-HIFME experience were the same as those described in Section 3.2.2.1.

I recruited sixty-six participants (33 in each instructions group; 23 men and 43 women: M_{age} = 21.2 years old, *SD* = 2.0 years old) from the local and student populations in St Andrews. I removed data from four participants, because they reported previous HIFME experience. I also removed data from one participant for the Rounds analysis (for details on the Rounds workout, see Section 5.2.2.2), because the participant misunderstood the workout structure, and thus made predictions that were not based on the actual workout. Additionally, I removed data from an outlier (a woman in the impulsive group) for the Rounds analysis, and data from another outlier (a woman in the impulsive group—not the same outlier as in the Rounds workout) for the AMRAP workout analysis (for details on the AMRAP workout, see Section 5.2.2.2). I classified participants as outliers if the absolute value of their absolute accuracy *z*-score was larger than three.

For calibration analyses in the Rounds workout, I analysed data from sixty participants (29 in the impulsive group and 31 in the strategic group; 22 men and 38 women; M_{age} = 20.9 years old, *SD* = 1.6 years old). For calibration analyses in the AMRAP workout, I analysed data from sixty-one participants (30 in the impulsive group and 31 in the strategic group; 22 men and 39 women; M_{age} = 20.9 years old, *SD* = 1.5 years old). For data checks (see Section 5.2.3.1), I analysed data where I had excluded outliers from both workouts, leading to a sample size of 59 participants (28 in the impulsive group and 31 in the strategic group; 22 men and 37 women; M_{age} = 20.9 years old, *SD* = 1.6 years old).

The study received ethical approval from the University of St Andrews School of Psychology & Neuroscience Ethics Committee (Ethics approval code: PS14081; see Appendix 8.1.6). All participants were compensated at a rate of £5/hour.

5.2.2.2 HIFME workouts

Rounds workout. In the first workout, I asked participants to complete three rounds of 5 push-ups (Fig. 5.1), 10 sit-ups (Fig. 5.2), and 20 alternating lunges (i.e. lunges where athletes had to alternate legs for every repetition—10 lunges for each leg; Fig. 5.3) as quickly as possible. I measured performance using finish time, i.e. the time it took for a participant finish all three rounds. For their predictions, I asked participants to indicate how many minutes and seconds it would take them to complete all three rounds. I measured both performance and predictions in terms of seconds (I converted minutes to seconds for all analyses).

Figure 5.1

Demonstration of push-ups.

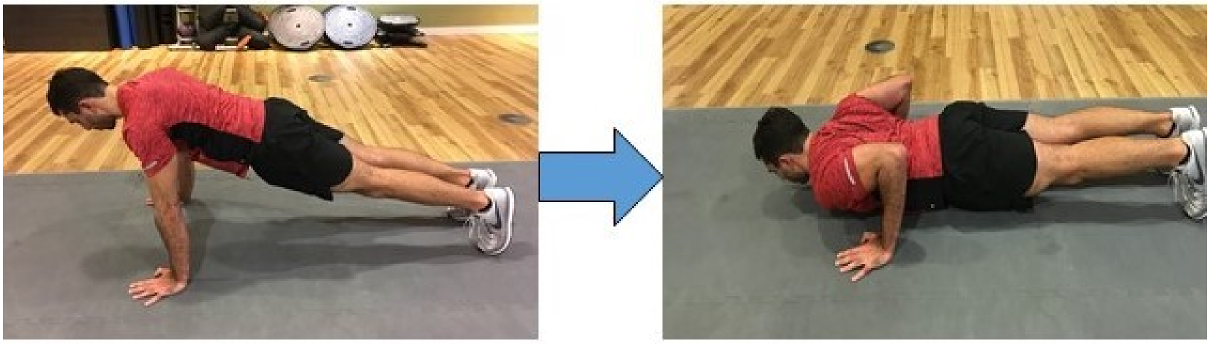


Figure 5.2

Demonstration of sit-ups.

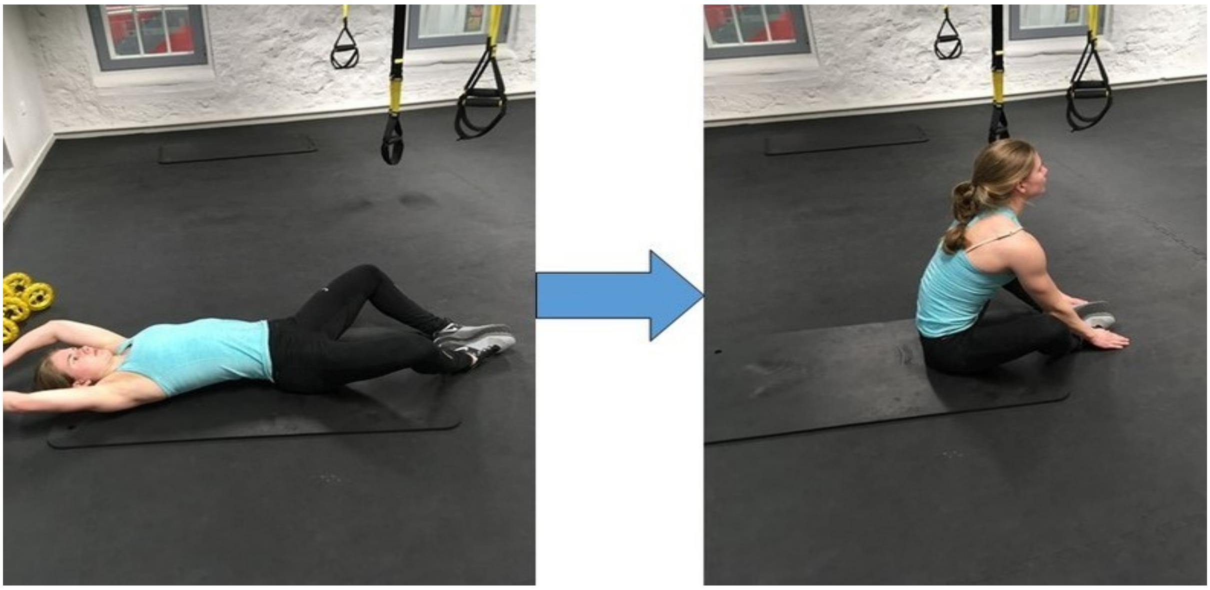
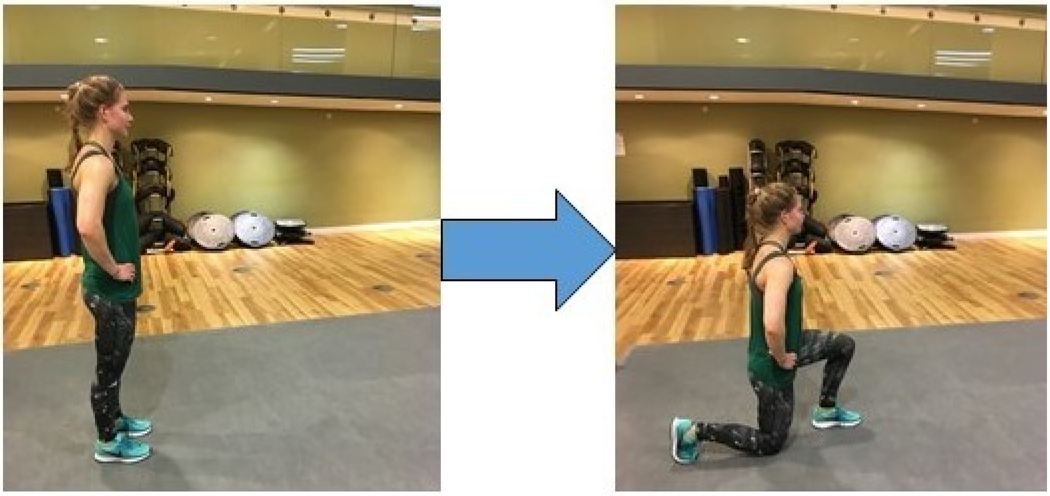


Figure 5.3

Demonstration of lunges.



As Many Rounds As Possible (AMRAP) workout. In the second workout, I asked participants to complete as many rounds of 5 burpees (Fig. 5.4), 10 back extensions (Fig. 5.5), and 20 mountain climbers (alternating between legs – 10 mountain climbers per leg; Fig. 5.6) as possible in five minutes. I measured performance in terms of the number of repetitions participants completed in the five minutes. I asked participants to give their predictions in terms of the number of rounds and repetitions they expected to complete. I measured both performance and predictions in terms of repetitions, and not rounds (for example, a prediction of five rounds equalled to 175 repetitions).

Figure 5.4

Demonstration of burpees

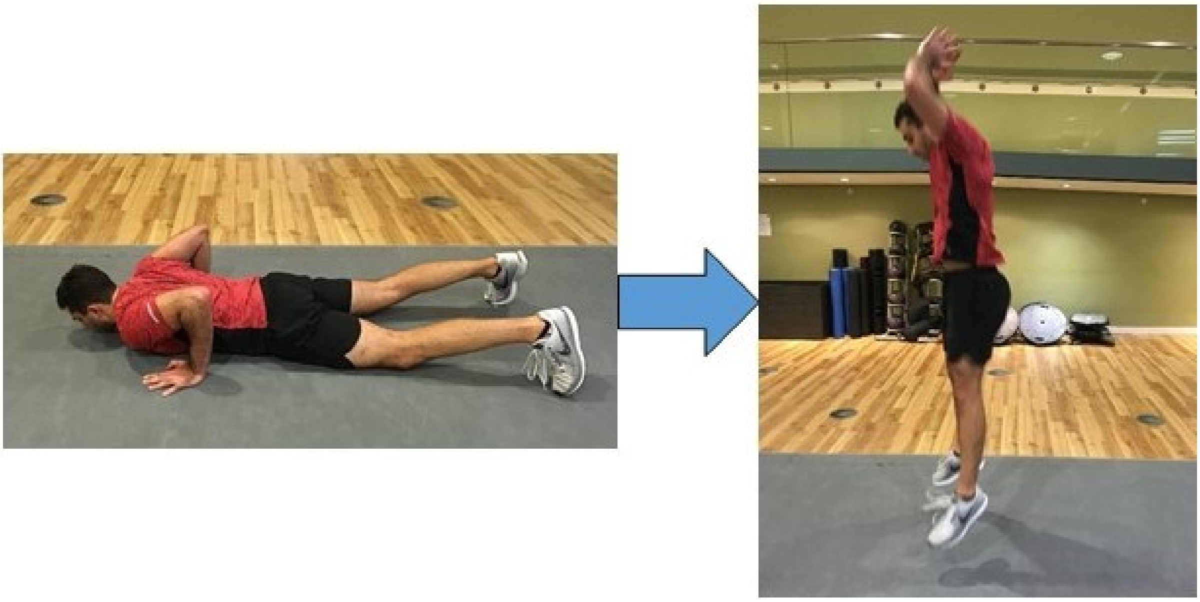


Figure 5.5

Demonstration of back extensions.

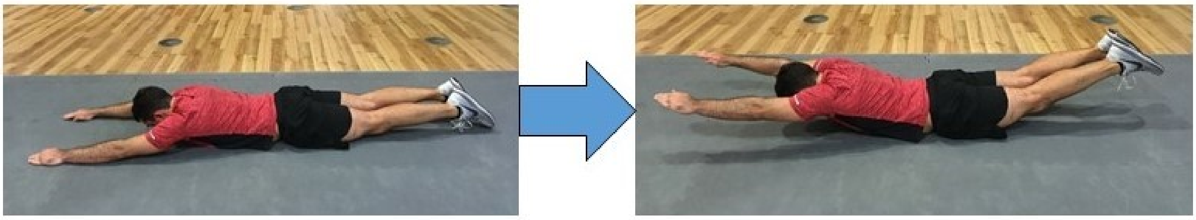
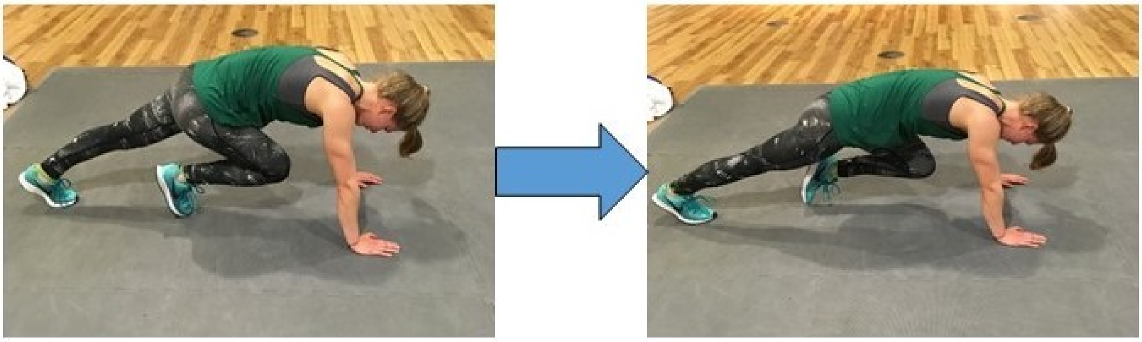


Figure 5.6

Demonstration of mountain climbers.



5.2.2.3 Prediction instructions

Prior to the first workout, I gave all participants the same instructions on how to make their performance predictions. I asked them to produce realistic estimates of how they thought they were going to perform, as opposed to how they hoped to perform. In the second workout, I once again asked all participants to make realistic estimates of their performance, but then presented each group with different prediction instructions. The instructions for the impulsive group were the following: “I want you to give me a prediction based on your gut feeling—do not think about it at all. Simply give me the first prediction that comes to your mind, without engaging in any strategic thinking.” I also advised any participants who attempted to engage in strategic thinking to avoid doing so. The strategic group received the following instructions: “I want you to think about the prediction you give me and be strategic about it. One good strategy would be to estimate the time you think you require to complete a round, adjust the time for fatigue based on the time cap of 5 minutes, and then think of how many such rounds you can complete in 5 minutes.” After either instruction condition, I asked participants to verify that their predictions were impulsive or strategic.

5.2.2.4 Design

The aim of Study 7 was to examine the effects of prediction guidance on HIFME calibration in participants who did not have previous HIFME experience. I compared calibration between the two groups in both workouts to ensure that there were no baseline differences in the first workout, and to test whether the manipulation of instructions affected calibration in the second workout. I assessed calibration using bias and absolute accuracy percentages relative to performance (formulae for these calculations in HIFME can be found in Section 3.2.2.4). Using percentages allowed me to control for the effects of performance value variation on calibration.

5.2.2.5 Procedure

I emailed participants with a Participant Information Sheet that contained details about the study, and a PAR-Q health eligibility form (see Appendix 8.5.2) to ensure that no previous health conditions would be exacerbated during the workouts. I assigned eligible participants a study slot, during which I asked them to provide their informed consent, and indicate their exercise experience verbally to ensure they had no HIFME background. During a 10-minute warm-up period prior to the HIFME workouts, I demonstrated the exercise standards and technique to participants, who then completed the exercises themselves. I also asked participants to indicate whether they had been familiar with each exercise before the study to ensure that neither group was more familiar with the exercises than the other. Following the warm-up period, I presented participants with the first workout and asked them to provide their performance predictions. As described in Section 5.2.2.3, the prediction instructions were the same across groups for the first workout. After participants completed the first workout, they had a 5-minute break to rest. Two minutes into the break, I presented participants with the second workout, for which I instructed them to make their predictions as described in Section 5.2.2.3 based on their randomly assigned group. Following the conclusion of the second workout, I debriefed participants, and the study concluded.

5.2.3 Results

5.2.3.1 Data checks

Exercise familiarity. To examine differences in exercise familiarity collapsed across the two workouts between the impulsive and the strategic groups, I conducted a Mann-Whitney U test. I used a non-parametric test, as familiarity scores were not normally distributed. The test indicated that participants in the two groups reported similar familiarity with the exercises they completed across the two workouts (Impulsive: $M = 5.57$, $SD = 0.63$; Strategic: $M = 5.55$, $SD = 0.77$), $U = 447.00$, $p = .811$. The lack of difference in familiarity with workout movements between the two groups suggests that participants in the two groups were similarly experienced with HIFME, and that exercise familiarity was unlikely to affect subsequent analyses.

Exercise experience. To examine potential differences in general exercise experience between the two groups, I asked participants to indicate the number of years during which they had engaged in exercise (any modality). I conducted an independent samples t-test, which showed that participants in the impulsive instructions group ($M = 7.8$ years, $SD = 4.5$ years) had similar exercise experience with the strategic instructions group ($M = 7.5$ years, $SD = 4.1$ years), $t_{(57)} = 0.27$, $p = .786$, $d = 0.07$. Therefore, the random allocation of participants into the two groups was successful in ensuring that participants of either group had similar exercise experience and familiarity with the exercise movements comprising the two workouts. This means that participants should exhibit similar calibration in the Rounds workout, and that any differences in calibration between the two groups in the AMRAP workout should result from the experimental manipulation of prediction instructions, as opposed to differences in exercise experience.

Gender distribution. To ensure that the distribution of male and female participants was similar across the two groups, I conducted a chi-squared test. The test showed that there was no significant difference in gender distribution across the two groups, $\chi_{(1)} = 0.09$, $p = .763$. There was thus a similar number of male and female participants in each group.

5.2.3.2 Performance & Predictions

To examine group differences in performance and predictions across the two workouts, I conducted independent samples t-tests.

Rounds. There was no significant difference in Rounds performance between the impulsive and the strategic groups, $t_{(58)} = -0.93$, $p = .356$, $d = -0.24$ (Table 5.1), indicating a lack of difference in athletic capacity between the two groups. In the same vein, participants in the impulsive group made similar performance estimates as participants in the strategic group, $t_{(58)} = -0.69$, $p = .495$, $d = -0.18$ (Table 5.1). Overall, findings on Rounds performance and predictions, along with findings on exercise experience and familiarity, suggest that the random placement of participants in the two groups was successful, as there were no baseline differences in performance, predictions, experience, and exercise familiarity between the two groups.

AMRAP. Similar to the Rounds workout, there were no group differences in AMRAP performance between the impulsive and the strategic groups, $t_{(59)} = 0.66$, $p = .515$, $d = 0.17$ (Table 5.1). This means that the manipulation of prediction instructions did not affect workout performance. Interestingly, participants in the strategic group exhibited a non-significant tendency to predict they would complete fewer repetitions in the AMRAP workout than participants in the impulsive group, $t_{(59)} = 1.80$, $p = .077$, $d = 0.46$ (Table 5.1). This suggests that strategic instructions had a minor effect in eliciting predictions that were more conservative than instructions for participants to be impulsive in their predictions. This difference could lead to differences in the AMRAP bias analysis presented below.

Table 5.1

Descriptive statistics for the two instructions groups.

	Impulsive Group		Strategic Group	
Outcome Variable	***Mean***	***SD***	***Mean***	***SD***
Rounds Performance	215 s	47 s	227 s	47 s
Rounds Predictions	242 s	104 s	262 s	117 s
AMRAP Performance	188 reps	43 reps	180 reps	54 reps
AMRAP Predictions	185 reps	90 reps	149 reps	65 reps
Rounds Bias	12.01%	40.28%	15.57%	43.09%
Rounds Absolute Accuracy	34.31%	23.50%	34.15%	30.04%
AMRAP Bias	-0.55%	43.18%	-18.08%	18.01%
AMRAP Absolute Accuracy	36.16%	22.62%	21.98%	12.76%

Note. The table provides information on the means and standard deviations (SD) of participants who received impulsive instructions and participants who received strategic instructions for performance, predictions, and calibration across workouts.

5.2.3.3 Calibration

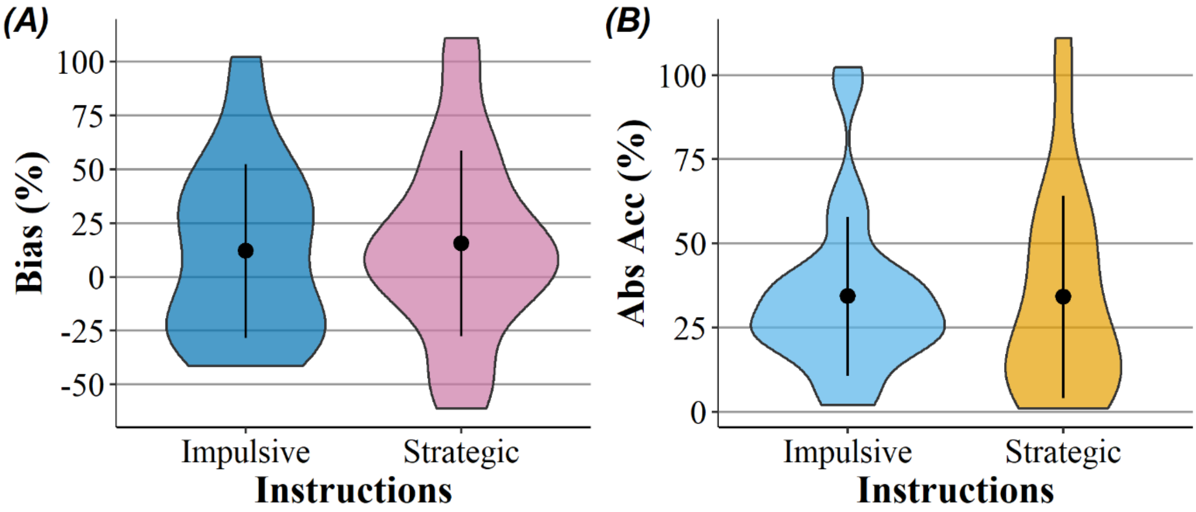
To examine group differences in calibration for the Rounds and the AMRAP workouts, I conducted independent samples t-tests. This statistical analysis was different to the methods I used to examine calibration in previous chapters, where I conducted multiple linear regressions with and without the inclusion of performance as a predictor. This is because in previous chapters I examined the associations of numerous factors with calibration, which often exhibited relationships with performance. However, in the present study, I only focused on the effects of prediction instructions on calibration, which, as illustrated above, did not influence performance in either workout. Therefore, I did not anticipate that the inclusion of performance in the analyses would influence the effects of prediction instructions on calibration, thus excluding it from the present calibration analyses.

Rounds. Positive bias scores in the Rounds workout indicate underconfidence, whereas negative scores indicate overconfidence. Overall, participants were significantly underconfident (M = 13.85%, SD = 41.44%) in their predictions for the Rounds workout, $t_{(59)} = 2.59$, $p = .012$, $d = 0.33$. The t-test comparing Rounds bias between the two instructions groups did not find a significant difference between them, $t_{(58)} = -0.33$, $p = .743$, $d = -0.09$. Participants who received impulsive instructions were as likely to be underconfident as participants who received strategic instructions (Table 5.1; Fig. 5.7A). In the same vein, participants in the impulsive group were similarly precise as participants in the strategic group, as there was no significant difference in absolute accuracy between them, $t_{(58)} = 0.02$,

p = .981, d = 0.01 (Table 5.1; Fig. 5.7B). These results suggest that there were no baseline differences in calibration between the two groups, mirroring exercise experience, performance, and prediction findings. This means that any potential differences in calibration between the two groups in the AMRAP workout should be attributed to the effects of instruction manipulations.

Figure 5.7

Violin plots illustrating the effects of instructions on bias and absolute accuracy in the Rounds workout.



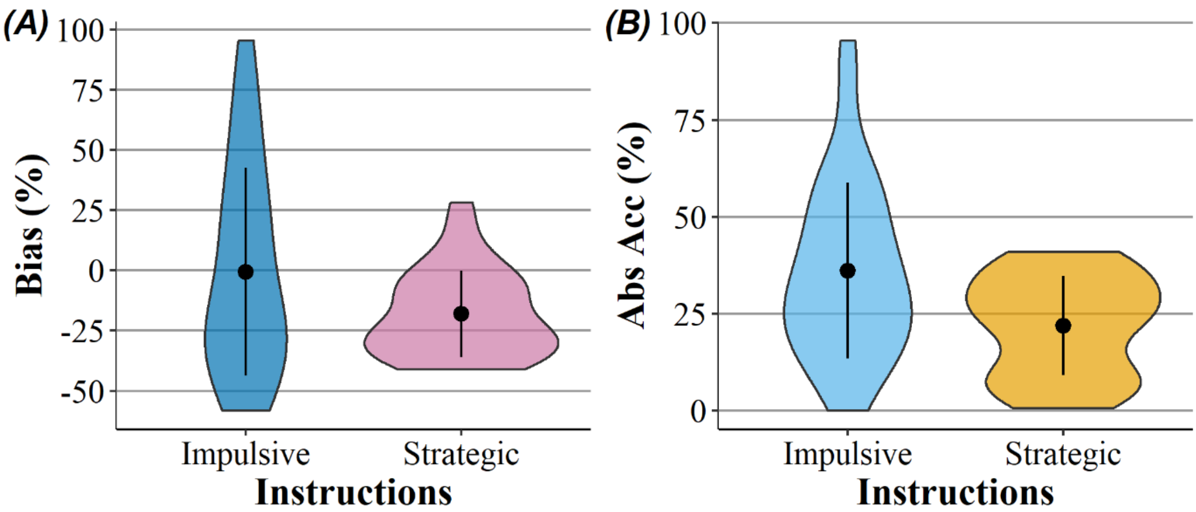
Note. Panel A shows bias differences in the Rounds workout between the two instruction groups. Panel B shows absolute accuracy differences in the Rounds workout between the two groups. The perimeter of each plot illustrates density, the central point represent the mean, and the vertical line represents +/- one standard deviation.

AMRAP. Positive bias scores in the AMRAP workout indicate overconfidence, whereas negative scores indicate underconfidence. As with the Rounds workout, participants were significantly underconfident (M = -9.46%, SD = 33.78%) in their predictions for the AMRAP workout, $t_{(60)} = -2.19$, p = .033, d = -0.28 . This illustrates that participants were consistently underconfident across the two workouts. Contrary to the Rounds workout, however, there was a significant difference in bias between the two instruction groups, $t_{(38.53)} = 2.06$, p = .047, d = 0.53—the Levene's test for equality of variances was significant, F = 22.54, p = .001, so I used Welch's t-test instead. Though participants in the impulsive group were unbiased, $t_{(29)} = -0.07$, p = .945, d = -0.01, participants in the strategic group were significantly underconfident, $t_{(30)} = -5.59$, p < .001, d = -1.00. (Table 5.1; Fig. 5.8A). This result thus indicated that strategic instructions led to underconfident predictions, whereas impulsive instructions led to unbiased predictions.

Nonetheless, the underconfidence and lack of bias observed in the strategic and impulsive groups respectively do not necessarily indicate that strategic predictions were generally more poorly calibrated than impulsive predictions, as individuals can make underconfident predictions that are also precise. Indeed, findings for absolute accuracy were not in line with those for bias. Participants who received strategic instructions were significantly more precise than participants who received impulsive instructions, $t_{(45.44)} = 3.00$, $p = .004$, $d = 0.77$ (Table 5.1; Fig. 5.8B)—the Levene's test for equality of variances was significant, $F = 6.84$, $p = .011$, so I used Welch's t-test instead. The present results suggest that, though the strategic instructions I used in Study 7 led to higher prediction underconfidence than the impulsive instructions, they also led to higher prediction precision compared to impulsive instructions. Overall, the extent to which athletes make strategic or impulsive predictions in a novel exercise modality appears to affect their calibration.

Figure 5.8

Violin plots illustrating the effects of instructions on bias and absolute accuracy in the AMRAP workout.



Note. Panel A shows the effects of instructions on AMRAP bias. Panel B shows the effects of instructions on AMRAP absolute accuracy. The perimeter of each plot illustrates density, the central point represent the mean, and the vertical line represents +/- one standard deviation.

5.2.4 Discussion of Study 7

The aim of Study 7 was to examine whether manipulating guidance on how to make performance predictions would affect HIFME calibration in participants without previous HIFME experience. I anticipated that participants who received metacognitive instructions to be strategic in their predictions by chunking their estimates in smaller components (i.e. AMRAP rounds) would be better calibrated (i.e. less biased and more precise) than participants who received instructions to be impulsive in their predictions by providing the first estimate that came to their minds.

The results partially supported my prediction. Instructing participants to metacognitively strategize when making predictions for the AMRAP workout led to differences in both bias and absolute accuracy between the two groups. Surprisingly, participants who made impulsive predictions were unbiased, whereas participants who made strategic predictions were underconfident. Thus, strategic instructions did not have the expected effect on bias, as they led to overly cautious predictions instead. This is an interesting finding, because it illustrates the importance of using prediction instructions that are appropriate for the athletic task examined. The AMRAP workout only lasted for five minutes, so it is likely that instructing participants to "adjust for fatigue" in their predictions led to them overestimating the fatigue they would experience during the workout, despite the instructions suggesting that this adjustment should be based on the time cap of five minutes. These instructions might have thus been more appropriate for endurance and long-lasting activities instead (Brick, Campbell, et al., 2016; Brick, MacIntyre, et al., 2016; Brick et al., 2015), where athletes would not be able to sustain maximal or near-maximal effort throughout the task's duration in the same way they could do in the present AMRAP workout. In that case, excluding the instruction to adjust for fatigue could have led to less biased strategic predictions. Interestingly, participants who did not engage in strategic thinking for their predictions did not exhibit any bias, indicating that strategic predictions can be less effective in producing low bias than impulsive predictions, if the strategy used is suboptimal.

Nonetheless, the bias findings alone cannot provide definitive evidence on strategic instructions leading to poorer overall calibration compared to impulsive instructions. It is possible for athletes to make underconfident predictions that are precise, i.e. overly conservative predictions that exhibit small deviations to performance, and unbiased predictions that are imprecise, i.e. predictions that are neither underconfident nor overconfident but exhibit large deviations to performance. Absolute accuracy findings were in line with such a pattern of results. Participants who made strategic predictions were more precise than participants who made impulsive predictions, supporting my prediction that engaging in metacognitive thinking when making predictions would lead to higher precision than making impulsive predictions.

Overall, participants who made impulsive predictions did not exhibit systematic bias in their estimates, but deviated from their performance more than participants who made strategic estimates, suggesting that impulsive predictions were more random in nature. In contrast, despite being underconfident in their performance estimates, participants who engaged in metacognitive thinking were more likely to make predictions that closely matched their performance. We can infer that these findings did not result from baseline group differences in performance, exercise experience, or calibration, as participants from either group reported similar exercise experience and familiarity, and exhibited similar performance and calibration in the baseline Rounds workout. Therefore, the present results illustrate that the way in which participants make predictions in a novel exercise modality can affect their subsequent calibration. This is an important finding, which highlights the necessity of providing guidance to new athletes on how they should be assessing their prospective performance.

An implication of Study 7 is that instructing participants to engage in strategic thinking based on metacognitive processes can lead to more precise performance assessments than asking them to make impulsive, non-strategic assessments. Furthermore, it appears that strategic instructions need to be tailored specifically to each workout type or exercise modality, as the same strategies might not be as effective in different situations (e.g. short versus long workouts). However, results from Study 7 have only shown that metacognitive instructions can facilitate calibration in a novel workout and might not generalise to situations where athletes are already familiar with an exercise modality, as I only recruited participants without previous HIFME experience. If we assume that experienced athletes typically engage in at least some strategic assessment of their task-specific performance, then it is possible that they will have acquired more metacognitive knowledge in relation to their task performance, as well as faster/better access to it. This could lead to them making impulsive predictions with similar calibration to their strategic predictions. I examine this possibility in Study 8.

5.3 STUDY 8 – INSTRUCTIONS & RUNNING CALIBRATION

5.3.1 Study specifics

In Study 8, I investigated the extent to which simply asking runners to make strategic or impulsive predictions would affect their running calibration. I recruited runners because running is a very popular sport (Andersen, 2020), and, given the small local and student population in St Andrews, it would have likely been more difficult to recruit athletes with experience in sports that are more specialised, and possibly less popular, than running (e.g. HIFME). Furthermore, running is a relatively straightforward activity since it involves athletes engaging in a singular movement pattern. Consequently, it is less complex and unpredictable than activities that include a wide range of movements and structures (e.g. HIFME; for discussion of exercise complexity, see Section 3.1). Because of this, minimal instructions could have a larger effect on calibration in running compared to other, more complex activities.

As discussed above, athletes with previous exercise experience might have more metacognitive knowledge regarding their performance, as well as better access to it, leading to impulsive and strategic predictions with similar calibration. Contrary to Study 7, where I split participants into an impulsive and a strategic group and used a Rounds workout to examine baseline differences in calibration, in Study 8, I asked the same participants to provide both impulsive and strategic predictions. I did this because, in Study 7, I wanted to first familiarise inexperienced participants with HIFME and the processes involved in assessing prospective HIFME performance before asking them to make impulsive or strategic HIFME predictions. If I asked participants to make predictions for the Rounds workout first, and then to make both impulsive and strategic predictions for the AMRAP workout, it is possible that they would experience fatigue and/or infer the study's aim and adjust their behaviour accordingly. However, since participants in Study 8 had previous experience with the athletic activity examined, I did not need to include a baseline workout to familiarise them with the task modality and the processes involved in predicting running performance. I thus asked the same participants to make both impulsive and strategic predictions instead, which allowed me to collect data from a larger sample size efficiently, and to increase the study's power by reducing variance related to individual differences.

Another difference between Studies 8 and 7 was the type of prediction guidance I used. The strategic instructions I used in Study 7 led to underconfidence, whereas impulsive predictions were unbiased. As I discussed above, this was likely the result of the strategic instructions being more appropriate for longer endurance, rather than shorter high-intensity, activities. To avoid this issue in Study 8, instead of providing participants with specific instructions on how to make their strategic predictions, I merely asked them to make a strategic prediction using any strategy they preferred (see Section 4.3.2.4 for instructions). A benefit of this minimal intervention is that it allowed me to observe whether the strategies that experienced athletes implement by themselves lead to better calibration than making impulsive predictions. This is an interesting topic to examine with important implications for athletes, as potential results could indicate whether simply asking athletes to be strategic and not impulsive in their predictions could have an effect in facilitating their performance awareness accuracy.

Overall, I anticipated that participants would be better calibrated when they made strategic rather than impulsive predictions. I predicted that simply asking participants to come up with their own strategies for their predictions would eliminate the strategic prediction underconfidence observed in Study 7, which I speculated to be the result of suboptimal strategic instructions. In line with Study 7, I predicted that performance estimates in the strategic condition would be more precise than estimates in the impulsive condition. Nonetheless, since participants had previous running experience and the intervention was minimal, I was also interested in the extent to which impulsive predictions could exhibit similar calibration to strategic predictions.

5.3.2 Methods

5.3.2.1 Participants

Participants in the present analysis came from the sample used in Study 5, which I presented in Chapter 4 (see Section 4.3.2.1). Sixty-seven runners (33 men and 34 women; M_{age} = 23.3 years old, *SD* = 5.5 years old) between the ages of 18 and 40 years old participated in the study, two of which did not finish the workout. Sixty-two participants completed the workout and gave both impulsive and strategic predictions (three participants indicated that their impulsive predictions were strategic, and were thus excluded from the analyses comparing impulsive and strategic predictions). There were two outliers in impulsive predictions (absolute values of *z*-scores were larger than three), and no outliers in strategic predictions. After removing outliers and participants who did not provide the intended predictions, I analysed data from sixty participants (29 men and 31 women; M_{age} = 23.2 years old, *SD* = 5.4 years old).

The study received ethical approval from the University of St Andrews School of Psychology & Neuroscience Ethics Committee (Ethics approval code: PS14429; see Appendix 8.1.5). All participants were compensated at a rate of £5/hour.

5.3.2.2 Running workout

Details on the running workout I used in the present study are the same as the ones used in Study 5 in Section 4.3.2.3.

5.3.2.3 Prediction instructions

Details on the instructions I provided participants for impulsive and strategic can be seen in Study 5 in Section 4.3.2.4.

5.3.2.4 Design

The aim of Study 8 was similar to that of Study 7, as I examined whether asking participants with previous running experience to make impulsive and strategic predictions would affect their running calibration. Contrary to Study 7, where I used a between subjects design, I implemented a within subjects design in Study 8 to increase its power in observing calibration differences between impulsive and strategic predictions. I assessed calibration using bias and absolute accuracy percentages relative to performance (formulae for these calculations in running can be found in Section 2.2.2.3). Using percentages allowed me to control for the effects of performance value variation on calibration. I could not counterbalance the order of prediction instructions, as asking for strategic instructions first would likely lead to participants making the same predictions across instruction conditions.

5.3.2.5 Procedure

Details on the procedure I used in Study 8 can be seen in Study 5 in Section 4.3.2.6.

5.3.3 Results

5.3.3.1 Performance & Predictions

Performance. In the present study, I used a within subjects design, so there were no performance differences between groups to examine. On average, it took participants 263 seconds (*SD* = 49 s) to complete the running distance of 1km.

Predictions. To examine differences between impulsive and strategic predictions for running performance, I conducted a paired sample t-test. The test showed there was no significant difference between the two prediction conditions, $t_{(59)} = 1.24$, $p = .220$, $d = 0.16$, despite participants showing a minor tendency to predict slower finish times in their impulsive predictions ($M = 276$ s, $SD = 80$ s) compared to their strategic predictions ($M = 267$ s, $SD = 61$ s). Therefore, participants were likely to make similar predictions across conditions.

5.3.3.2 Calibration

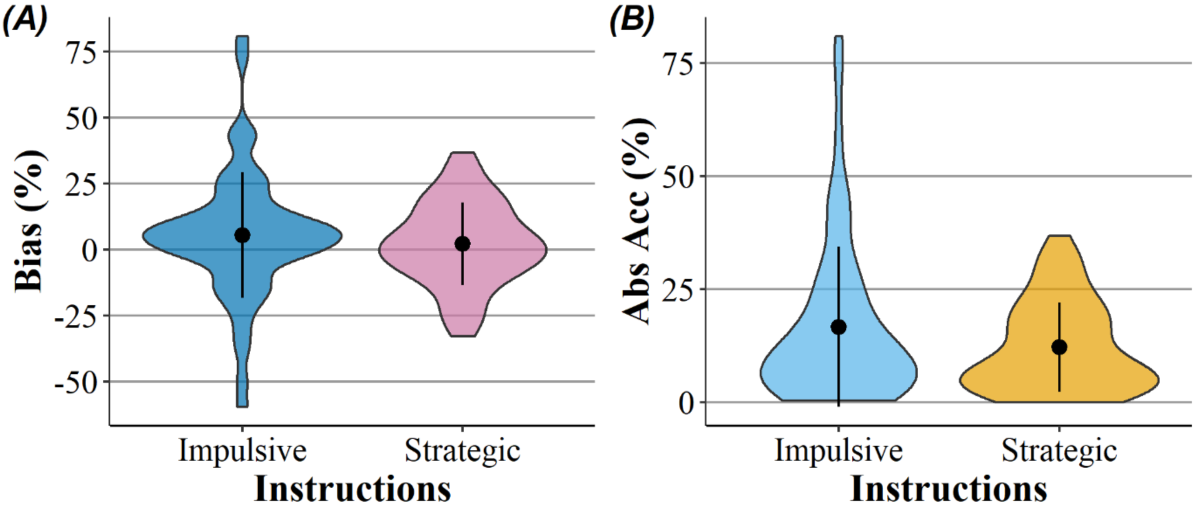
To examine the effects of instructions on calibration, I conducted repeated samples t-tests that compared bias and absolute accuracy between strategic and impulsive conditions.

Bias. Positive bias scores in the running workout in Study 8 indicate underconfidence, whereas negative scores indicate overconfidence. Overall, participants exhibited a non-significant tendency towards underconfidence in both their impulsive ($M = 5.50\%$, $SD = 23.78\%$), $t_{(59)} = 1.79$, $p = .078$, $d = 0.23$, and strategic predictions ($M = 2.22\%$, $SD = 15.60\%$), $t_{(59)} = 1.11$, $p = .274$, $d = 0.14$, though this tendency appeared to be stronger for the impulsive condition compared to the strategic condition. Bias comparisons between conditions produced similar results to the analysis of predictions. Though participants exhibited a tendency to be more underconfident in their impulsive predictions than their strategic predictions, this tendency was not significant, $t_{(59)} = 1.24$, $p = .222$, $d = 0.16$ (Fig. 5.9A). Therefore, instructing participants to make impulsive or strategic predictions did not have a significant effect on running bias.

Absolute accuracy. Contrary to bias findings, the analysis of running absolute accuracy exhibited a significant effect of prediction instructions on absolute accuracy, $t_{(59)} = 2.01$, $p = .049$, $d = 0.26$. Participants were significantly more precise in their strategic predictions ($M = 12.17\%$, $SD = 9.88\%$) than in their impulsive predictions ($M = 16.64\%$, $SD = 17.73\%$; Fig. 5.9B). This finding suggests that simply asking participants to make strategic predictions can lead to more precise performance estimates than asking them to make non-strategic, impulsive predictions.

Figure 5.9

Violin plots illustrating the effects of instructions on bias and absolute accuracy in the running workout.



Note. Panel A shows the effects of instructions on running bias. Panel B shows the effects of instructions on running absolute accuracy. The perimeter of each plot illustrates density, the central point represent the mean, and the vertical line represents +/- one standard deviation.

5.3.4 Discussion of Study 8

The aim of Study 8 was to examine whether asking athletes with previous running experience to make impulsive or strategic predictions would affect their calibration in a 1km running trial. I anticipated that bias findings would contrast those of Study 7 following changes in the strategic guidance used, eliminating the tendency of strategic predictions to be more underconfident than impulsive predictions. Furthermore, I expected that, similar to Study 7, participants in Study 8 would make strategic predictions that were more precise than their impulsive predictions, though I was also interested in the extent to which participants having previous running experience would affect this relationship.

Results from Study 8 were in line with my expectations. Contrary to Study 7, where participants who made strategic predictions were more underconfident than participants who made impulsive predictions, participants in Study 8 exhibited similar bias in their impulsive and strategic predictions. In fact, there was a numerical tendency for strategic predictions to be less underconfident than impulsive predictions. This finding highlights the importance of providing athletes with appropriate strategic instructions for their predictions, as suboptimal instructions can produce unexpected and undesired bias outcomes, such as the underconfidence observed in Study 7.

Replicating absolute accuracy findings from Study 7, participants were more precise in their strategic predictions than in their impulsive predictions. This illustrates that coaches and fitness instructors should encourage their athletes to engage in strategic thinking whenever they estimate their performance and plan their training or competition strategies, regardless of whether they have previous exercise experience or not. Neglecting this process and relying on non-strategic and impulsive assessments is likely to produce suboptimal calibration, and thus suboptimal performance outcomes (see Section 1.2.3.4). Interestingly, the effect of prediction instructions on absolute accuracy was smaller in Study 8 than Study 7. This could result from the use of minimal prediction guidance in Study 8 compared to Study 7 and/or the recruitment of experienced athletes, who likely had more metacognitive knowledge about their performance, as well as better access to it, than inexperienced athletes. In either case, the present study was important in demonstrating the need for athletes to engage in strategic thinking when estimating their prospective performance, and for coaches to encourage them to do so—even if that involves merely probing athletes to be strategic rather than impulsive in their predictions.

5.4 GENERAL DISCUSSION

In Chapter 5, I examined the effects of prediction guidance on HIFME and running calibration. In Study 7, I provided two groups of participants without previous HIFME experience with specific instructions on how to make either strategic or impulsive HIFME predictions. I anticipated that participants in the strategic group would exhibit better HIFME calibration (i.e. lower bias and higher precision) than participants in the impulsive group. In Study 8, I simply asked runners to produce impulsive and strategic predictions, without providing them with specific instructions on the strategy they should use. My predictions for Study 8 were similar to Study 7 in that I expected runners to be better calibrated in their running predictions following strategic, rather than impulsive, instructions.

Contrary to my expectations, participants who received strategic instructions in Study 7 were significantly more underconfident than participants who received impulsive instructions. It is not clear how these results relate to previous exercise research on the effects of metacognitive interventions on physical education calibration (Kolovelonis, Goudas, & Dermitzaki, 2012; Kolovelonis et al., 2013, 2020). Kolovelonis and colleagues (2020) suggested that participants who were underconfident and overconfident before a self-regulation intervention became less underconfident and overconfident respectively afterwards. However, the authors did not report inferential statistics for these analyses, so it is not possible to assess the reliability and strength of this finding. The authors also reported a non-significant tendency for most participants to be underconfident following the self-regulation intervention, whereas most participants were overconfident before it. In earlier studies conducted by the Kolovelonis lab (Kolovelonis, Goudas, & Dermitzaki, 2012; Kolovelonis et al., 2013), self-regulation interventions did not have a clear effect on bias relative to control conditions, though one of the experimental groups in the 2012 study exhibited post-training underconfidence—all other experimental groups were overconfident. Overall, the impact of metacognitive manipulations on exercise bias remains unclear, whilst findings from Study 7 are not sufficient to suggest that metacognitive manipulations lead to prediction underconfidence. This is because the strategic instructions implemented were likely a suboptimal choice for the short and high-intensity workout used, given that chunking and adjusting predictions for fatigue may be more appropriate for long and endurance activities (Brick, Campbell, et al., 2016; Brick, MacIntyre, et al., 2016; Brick et al., 2015).

Implementing strategic instructions that were more appropriate for the HIFME workout used could have produced a different effect on bias than the one observed.

To address this, in Study 8, I asked runners to use their own strategies in their performance estimates. I anticipated that allowing participants to use the strategy they believed to be most appropriate for the workout would lead to similar or better bias compared to the impulsive condition, contrasting findings from Study 7. In line with my prediction, there was no significant difference in bias between the strategic and impulsive conditions in Study 8. These results support the suggestion that it was the strategy used in the instructions in Study 7, which led to strategic produce underconfidence, rather than a general tendency of metacognitive instructions to induce underconfidence. It thus appears that the relationship between metacognitive manipulations and bias is not easy to define. This is likely because different types of interventions and instructions can affect how cautious participants are when they estimate their future performance. As illustrated in Study 7, unnecessary adjustments for fatigue in short workouts can lead to overly conservative predictions, which might be less likely to happen when participants do not receive instructions to make such adjustments. Overall, results from Chapter 5 illustrate that by manipulating metacognitive instructions, we can induce changes in prediction bias. However, they also highlight that these changes might not always be in the desired direction, and that we need to select task-appropriate instructions.

For absolute accuracy comparisons between strategic and impulsive prediction instructions, I anticipated that the former would lead to higher precision than the latter across studies. I was also interested in the extent to which participants having previous experience with the exercise modality used would eliminate or reduce precision differences between prediction instructions in Study 8. Results from both studies were in line with my expectations—participants exhibited higher precision in their performance estimates after receiving strategic instructions than after receiving impulsive instructions. Interestingly, the absolute value and effect size of this difference were larger in Study 7 (~14.2%) than in Study 8 (~4.4%). This suggests that previous running experience led to impulsive predictions that were more similar to strategic predictions in Study 8 than in Study 7, as participants did not have previous HIFME experience in the latter. The present results indicate that instructing participants to engage in strategic thinking is likely to lead to higher prediction precision than instructing them to produce impulsive predictions, regardless of previous exercise experience. This is in accordance with previous exercise and cognitive research (e.g. Gutierrez & Schraw, 2015; Gutierrez de Blume, 2017; Kolovelonis et al., 2020; Nietfeld et al., 2006), providing additional support for the effectiveness of various types of metacognitive interventions in improving calibration. Nonetheless, it also appears that previous task experience can affect the magnitude of this effectiveness—at least when we use minimal instructions as in Study 8.

The studies I presented in Chapter 5 have important theoretical and practical implications for exercise calibration. I was able to show that by simply instructing athletes to engage in metacognitive, rather than impulsive, thinking when producing metacognitive performance estimates, we can facilitate their calibration. This illustrates that even minimal interventions, such as prediction guidance, can be successful in inducing changes in calibration, as long as they are also effective in producing changes in relevant metacognitive processes and behaviour (Hacker et al., 2012; Kolovelonis et al., 2013; Stone, 2000). More practically, the present findings demonstrate that, to facilitate exercise calibration, coaches and fitness instructors need to ensure that athletes engage in strategic thinking when estimating their prospective performance. Otherwise, they run the risk of athletes producing impulsive predictions, which can have poor calibration. This is especially important for new and inexperienced athletes, who are more likely to exhibit higher precision differences between their impulsive and strategic predictions than experienced athletes. In such cases, it might be useful for coaches to attempt to facilitate athlete calibration by providing inexperienced athletes with specific instructions on the types of strategies they should implement—though future research should examine whether non-specific instructions can also facilitate calibration in inexperienced athletes—when they assess their prospective performance. Nonetheless, as illustrated by bias results in Study 7, not all strategies are effective in producing the desired outcome, and inappropriate strategic instructions can even have counterproductive effects on calibration (e.g. lead to underconfidence). Because of this, it is important to conduct further examinations of the types of strategic instructions best suited for different exercise modalities, intensities, and durations.

A potential issue with the two studies examined is whether participants made their predictions in accordance with the instructions they received. Though I asked participants to verify that they were strategic or impulsive when they predicted their performance, some might have indicated that they followed instructions without doing so. Such a discrepancy between reported and actual behaviour could have reduced statistical power, and thus the capacity of the two studies to observe differences between instructions. Nonetheless, I was still able to observe calibration differences between the two instruction types, suggesting that, even if some participants engaged in strategic thinking for their impulsive predictions and vice versa, they did so to an extent that did not eliminate the capacity of the two studies to observe calibration differences between conditions. Therefore, it appears that the manipulation of the instructions implemented was successful in eliciting the hypothesised patterns of behaviour across studies.

A second potential limitation is that instructing participants to produce impulsive predictions and comparing these to strategic predictions does not necessarily reflect actual performance-predicting behaviour. It is possible that when participants receive no instructions on how to make their predictions, their predictions range from being impulsive to being highly strategic. This means that the present results, which only compared strategic to impulsive instructions, could have overstated the benefits of strategic instructions in facilitating exercise calibration. Nonetheless, using impulsive predictions in the present studies was a deliberate choice aiming to maximise statistical power and to highlight the importance of engaging in strategic thinking as opposed to acting impulsively when predicting athletic performance. Examining whether athletes make strategic or impulsive predictions when they receive no instructions on how to make their predictions was thus outside the scope of Chapter 5. Future research should explore how athletes make their predictions when they receive no instructions, and how calibration in conditions without prediction guidance compares to calibration in impulsive and strategic conditions instead.

5.5 CONCLUSION

Chapter 5 provided experimental evidence on the effects of manipulating prediction guidance on exercise calibration. Participants who received instructions to engage in metacognitive thinking when assessing their prospective HIFME and running performance were more precise than participants who received instructions to produce impulsive, non-strategic, predictions. This result highlights the importance of ensuring that athletes take advantage of their metacognitive skills and knowledge when estimating performance, regardless of whether they have previous experience with an exercise modality. Bias findings also highlighted the importance of selecting appropriate strategic instructions, as instructions in Study 7 led to overly conservative predictions, whereas impulsive predictions were unbiased. After using different strategic predictions in Study 8, there were no longer bias differences between the two conditions, indicating that we need to tailor strategic instructions to the characteristics of each exercise task (e.g. type, duration, and intensity). Overall, the present results illustrate the importance of encouraging metacognitive and strategic thinking in performance predictions using appropriate prediction guidance. To my knowledge, the present studies were the first to examine the impact of prediction guidance on exercise calibration. It is thus important to build on their findings and further investigate the effects of metacognitive manipulations of prediction guidance on calibration in order to increase our understanding of how to optimise exercise calibration.

CHAPTER 6: THESIS DISCUSSION

6.1 SUMMARY OF RESEARCH PROJECT

In the present thesis, I examined exercise calibration using running and high-intensity functional movement exercise (HIFME). Specifically, I focused on whether demographic factors, self-reports of exercise metacognition, and cognitive calibration are effective in predicting athlete calibration. I also tested the extent to which we can use metacognitive manipulations of prediction guidance to optimise exercise calibration. To explore these associations, I presented eight studies in Chapters 2, 3, 4, and 5. I outline the purpose of each study and the methodologies I used below, before presenting the findings in the next section.

In Chapter 2 (Table 6.1), I conducted two observational studies to investigate the relationships between demographic factors and running calibration. For these studies, I collected demographic and prediction data from runners before they participated in a running race (10km in Study 1; half marathon in Study 2). Using these data, I tested the extent to which the demographic factors of expertise, experience, age, and gender were effective in predicting running calibration. I also conducted an experimental manipulation on the type of predictions participants made by asking them to produce realistic (i.e. finish time they most likely expected to achieve) and goal (i.e. the finish time they hoped to achieve—their goal time) predictions separately. This allowed me to control for prediction type in calibration analyses, ensuring that participants did not differ in the extent to which they indicated their expected or goal finish times, and to investigate whether different prediction types lead to discrepancies in running calibration patterns. In Study 2, I also explored whether prediction lead, i.e. the number of days before the race when participants made their predictions, would be associated with running calibration.

In Chapter 3 (Table 6.1), I presented data from one observational study (Study 3), in which I also examined the relationships between demographic factors and exercise calibration. Unlike Chapter 2, where I examined these relationships in running, I used HIFME in Chapter 3, because it is a more complex and unpredictable exercise modality than running, as it involves a wider range of exercise movements and structures, which could make it harder for athletes to predict their prospective performance. Furthermore, there have been no previous investigations of calibration in HIFME, so this was a novel research area to explore in the field of exercise calibration. The demographic factors I examined were expertise, experience, and gender. Contrary to the studies in Chapter 2, I did not examine age and prediction type influence on calibration. Instead, I controlled for both factors by only recruiting individuals between the ages of 18 and 40 years old, and by explicitly instructing participants to produce realistic and not goal predictions. To measure HIFME calibration, I used two workouts that had different structures, because HIFME typically consists of a wide range of exercise movements and workout structures. Overall, Chapters 2 and 3 explored the relationships between demographic factors and exercise calibration, whilst controlling for prediction type. Results from these two chapters had important implications in understanding whether we can use athlete demographic information, which is easy and quick to collect, to make inferences about their bias and precision tendencies.

Table 6.1

Study designs in the thesis.

Chapter	Study	Number of participants	Research design	Key independent variables	Key dependent variables
Chapter 2	Study 1	189-199	Observational & Experimental	Expertise, Experience, Age, Gender, & Prediction type	Prediction bias & absolute accuracy
	Study 2	303-309	Observational & Experimental	Expertise, Experience, Age, Gender, Prediction lead, & Prediction type	
Chapter 3	Study 3	50-56	Observational	Expertise, Experience, & Gender	
Chapter 4	Study 4	44-54	Observational	MAIE Knowledge & Regulation	
	Study 5	58-63	Observational	MAIR Knowledge & Regulation	
	Study 6	52-55	Observational	Memory recognition bias & absolute accuracy	
Chapter 5	Study 7	59-61	Experimental	Prediction guidance	
	Study 8	61	Experimental	Prediction guidance	

Note. The table provides information on the designs of each study in the thesis. The first column indicates the chapter in which each study can be found. Some studies had multiple parts, which I presented in different chapters. In these cases, I present on the table the design of each study part in its relevant chapter. The second column presents the studies in the order they appeared in each chapter. The third column provides information on sample size for each study. For studies with different numbers of participants in different analyses (e.g. because of outliers), I present ranges of sample sizes. The fourth column indicates whether a study had an observational or an experimental design. The fifth column lists the key independent variables examined in each study, and the sixth column lists the key dependent variables.

In Chapters 4 and 5, instead of focusing on demographic factors, I explored the relationships between metacognitive behaviour and exercise calibration. More specifically, in Chapter 4 (Table 6.1), I presented data from three observational studies that explored whether self-reports of exercise metacognition and cognitive calibration are effective in informing us about exercise calibration. In Study 4, I tested whether self-reports of general exercise metacognition would predict HIFME calibration in participants without previous HIFME experience. In Study 5, I examined the same relationship using running-specific metacognition self-reports and running calibration instead. Participants in Study 5 had previous running experience. In Study 6, I collected calibration data using a memory recognition task, and tested whether cognitive calibration correlated with HIFME calibration. Findings from Chapter 4 had important implications for exercise calibration, as they indicated whether we could use self-reports of exercise metacognition and calibration metrics from non-exercise domains to assess athlete calibration in a time-efficient manner, without requiring athletes to first engage in an exercise task.

Finally, the two experimental studies I presented in Chapter 5 examined the effects of metacognitive manipulations on exercise calibration (Table 6.1). The metacognitive manipulation I implemented was prediction guidance, i.e. I provided participants with metacognitive strategic and non-metacognitive impulsive instructions on how they should predict their performance. In Study 7, I recruited participants without previous HIFME experience to investigate whether strategic instructions would lead to better calibration than impulsive instructions in an unfamiliar HIFME workout. For strategic predictions, I provided participants with detailed instructions, which described a specific metacognitive strategy on how to predict their performance. In Study 8, I recruited participants with previous running experience and used a minimal prediction guidance intervention, i.e. simply instructing participants to be strategic or impulsive in their predictions, to examine differences in running calibration between strategic and impulsive predictions. These studies were integral in better understanding the types of metacognitive manipulations that can optimise exercise calibration, as well as the extent to which coaches and fitness instructors could facilitate exercise calibration in experienced and inexperienced athletes simply by instructing them to be strategic in their predictions.

6.2 SUMMARY OF FINDINGS

The studies I presented in the thesis produced important results regarding exercise calibration. I summarise the main predictions and findings of each chapter below.

In Chapter 2, I expected that runners with higher expertise and experience would be better calibrated than runners with lower expertise and experience respectively. I also anticipated that older runners would be better calibrated than younger runners—at least before controlling for other experience factors—and that female runners would be less overconfident (or more underconfident) than male runners in their performance predictions. Furthermore, I predicted participants to be more overconfident in their goal predictions compared to their realistic predictions. Finally, I only examined prediction lead in Study 2, where I hypothesized that participants would be better calibrated when they made their predictions closer to the time of the race than earlier in advance.

Results from the two studies were partially in line with my predictions. As expected, participants with high expertise, i.e. who were faster to finish the race, were more precise and more likely to be underconfident, unbiased, or less overconfident than participants with low expertise. Surprisingly, experience factors did not exhibit a clear and reliable association with calibration. For example, though training volume and months of running experience exhibited a tendency to predict higher precision in Studies 1 and 2 respectively, accounting for performance reduced or eliminated these tendencies. Furthermore, the relationships between experience factors and bias were not consistent across studies. These results suggest that, though experience factors can play a role in running calibration, their role is not as clear as expected. Conversely, older age was associated with higher precision in goal predictions across studies, even after accounting for other experience factors. This suggests that older age has a positive contribution to running calibration that is independent to other experience factors. Interestingly, older runners also tended to be more underconfident (or less overconfident) across studies. In line with my expectations, female runners were more likely to be underconfident or less overconfident than male runners across studies, but only after accounting for finish time variance.

Prediction type results in Study 2 were partly in line with my hypothesis, as runners were overconfident in their goal predictions, but underconfident in their realistic predictions. More surprisingly, runners in Study 1 were underconfident in their realistic predictions (more so than runners were in Study 2), and unbiased in their goal predictions. Despite differences in bias directions between the two studies, results were consistent in exhibiting a difference in bias between prediction types across studies. Finally, as I anticipated for prediction lead, participants exhibited higher precision in their performance estimates when they made their predictions closer to the time of the race than earlier in advance. Overall, Chapter 2 findings provided intriguing information on the ways in which demographic factors, prediction type, and prediction lead relate to running calibration.

In Study 3 in Chapter 3, I predicted that better performers (i.e. with higher expertise), athletes with a HIFME background, and athletes with more HIFME experience would be better calibrated than poorer performers, athletes without HIFME background, and athletes with less HIFME experience respectively. Additionally, I predicted that male participants would be more overconfident/less underconfident than female participants. These predictions were concordant with predictions in Chapter 2. In line with my expectations, participants with a HIFME background were more precise in their performance predictions than participants without a HIFME background. However, this finding was only present in one of the two workouts examined. Surprisingly, expertise, gender, and HIFME experience factors in participants with a HIFME background contrasted my predictions and findings from Chapter 2, as they did not exhibit associations with calibration in either HIFME workout. This was a surprising finding since I anticipated demographic factors to exhibit consistent relationships with calibration across different exercise modalities. Though it is not clear why I observed discrepancies in findings between Chapters 2 and 3, present results suggest that the associations between demographic factors and with calibration are not consistent across different exercise modalities.

In Chapter 4, I expected that running-specific self-reports of metacognition would predict running calibration in runners with previous running experience in Study 5. In Study 4, I used an exercise-general questionnaire, and I only recruited participants without HIFME experience, so I did not make a specific prediction on whether metacognition scores would predict HIFME calibration. In Study 6, I predicted that calibration in a memory recognition task would correlate with calibration in two HIFME workouts.

Results from Studies 4, 5, and 6 in Chapter 4 were not consistent with my predictions. Scores on metacognition self-reports in Studies 4 and 5 did not predict exercise calibration, regardless of whether the questionnaires targeted general or task-specific exercise metacognition. Previous exercise experience did not play a role in these results either. In the same vein, cognitive calibration did not correlate with HIFME calibration in Study 6. This result was in line with Study 3, where calibration from one HIFME workout did not correlate with calibration from the other. However, the dissociation between the memory recognition task and the HIFME workouts in Study 6 was more prominent than the dissociation between HIFME workouts in Study 3. This is because there were correlations for performance predictions and performance between the two HIFME workouts in Study 3, but there were no such associations for metacognitive judgments and performance between the memory recognition task and the HIFME workouts in Study 6. Overall, results from Chapter 4 demonstrate a dissociation of offline measures of exercise metacognition and cognitive calibration with exercise calibration. Instead, it appears that the demographic factors examined in Chapters 2 and 3 are more informative regarding athlete calibration than the measures examined in Chapter 4, and should thus be preferred.

In Chapter 5, I predicted that instructing participants to produce strategic estimates of their performance would lead to better calibration than instructing them to produce impulsive estimates. In accordance with my prediction, strategic predictions led to higher prediction precision across studies, regardless of whether participants received specific or general strategic instructions, and whether they had previous experience with the exercise modality examined. The difference in precision between strategic and impulsive predictions was higher in Study 7, where participants receive specific strategic instructions and did not have previous HIFME experience, than in Study 8, where participants received minimal and general strategic instructions and had previous running experience. Surprisingly, participants who received strategic instructions were underconfident in their predictions in Study 7, whereas participants who produced impulsive predictions were unbiased. This was likely the result of the strategy described in the instructions being suboptimal in Study 7, as there was no difference in bias between the strategic and impulsive conditions in Study 8, where strategic instructions did not refer to a specific strategy. These results highlight our capacity to improve calibration by instructing athletes to produce strategic, rather than impulsive, predictions—though it is important to ensure that the strategies implemented are appropriate. Interestingly, Chapter 5 prediction guidance findings matched prediction type findings in Chapter 2 in that different prediction processes led to differences in calibration in both chapters, suggesting that how we ask athletes to make their performance predictions has important implications for calibration.

6.3 GENERAL DISCUSSION

Results from the present thesis were important in expanding exercise calibration research by furthering our understanding of the ways in which we can predict and optimise exercise calibration. In the following section, I discuss the implications of Chapters 2, 3, 4, and 5 findings in relation to relevant calibration literature.

Chapter 2 results on running expertise supported previous cognitive and exercise evidence on the association of high expertise with high precision and slight underconfidence, and low expertise with low precision and overconfidence (e.g. Kolovelonis, 2019; Kolovelonis & Goudas, 2018; Krawczyk & Wilamowski, 2016; Kruger & Dunning, 1999; Schlösser et al., 2013). In contrast, I did not observe a significant association between performance and HIFME calibration in Chapter 3—though not discussed explicitly, there were similar non-significant findings in Chapter 4. Chapter 3 results were instead in line with a small number of exercise studies that have observed a minor and inconsistent role of expertise in exercise calibration (Fogarty & Else, 2005; Fogarty & Ross, 2007). Overall, Chapter 2 findings point towards a positive association between expertise (i.e. performance) and exercise calibration, whereas we could attribute the lack of significant findings in the other chapters to low statistical power resulting from insufficient sample sizes.

Results on the relationship between experience and exercise calibration were consistent across Chapters 2 and 3. In accordance with previous calibration and pacing research (e.g. Deaner et al., 2014; Kolovelonis, 2019; Liverakos et al., 2018; Swain et al., 2019), experience markers in the two chapters exhibited positive associations with calibration, which were, however, minor and/or inconsistent. These results thus suggest that we should be cautious when we use experience indicators to predict exercise calibration, and avoid relying heavily on individual indicators. We should instead aim to use a wide range of experience markers that, taken together, can provide more detailed and reliable information regarding an athlete's calibration.

Interestingly, findings from Chapter 2 also indicated that, regardless of years of running experience, age could serve as an experience factor, as older runners were more precise in their goal predictions and more likely to be underconfident than younger runners. These results contrasted cognitive evidence of age-related decline in calibration (e.g. Cauvin et al., 2019; Soderstrom et al., 2012), and instead supported suggestions of age-related experience having a positive contribution to calibration in naturalistic settings (Cauvin et al., 2019; Devolder et al., 1990). They were also in line with pacing studies that have exhibited a relationship between older age and lower running pace slowing (e.g. Deaner et al., 2014; March et al., 2011; Trubee et al., 2014). A potential way in which age contributes to running experience could be that older runners have to engage in extensive metacognitive behaviour to account for age-related physical decline, leading to high performance awareness. Consequently, though the present age findings merit further investigation, they suggest that we should account for age along with other experience factors when we assess athlete calibration.

Chapters 2 and 3 produced inconsistent findings on gender differences in bias. In Chapter 2, female runners were more likely to be more underconfident or less overconfident than male runners after adjusting for finish time variance, in line with other running calibration and pacing studies (Deaner et al., 2016, 2014; Deaner & Lowen, 2016; Hubble & Zhao, 2016; Krawczyk & Wilamowski, 2016, 2018; March et al., 2011; Trubee et al., 2014). In contrast, there were no gender differences in HIFME bias between male and female participants in Chapter 3, mirroring physical education research (Kolovelonis, 2019; Kolovelonis & Goudas, 2018; Kolovelonis, Goudas, & Dermitzaki, 2012). The discrepancies in gender findings between the two chapters and exercise literature suggest that gender does not necessarily have a uniform influence on bias across different exercise modalities. Nonetheless, it is not clear why there were discrepancies in gender differences in bias across different exercise modalities in Chapters 2 and 3.

Overall, findings from Chapters 2 and 3 highlight the difficulties of relying on demographic factors to produce comprehensive assessments of athlete calibration. Though expertise, experience, age, and gender all showed associations with exercise calibration, these associations were often inconsistent across exercise modalities, suggesting that we need to also account for exercise modality when we use demographic factors to assess athlete calibration. Furthermore, most factors exhibited minor associations with calibration, highlighting the need to use large sample sizes to observe these relationships reliably. The weak contributions of individual factors to calibration also indicate that we need to account for numerous demographic factors before we can make reliable predictions regarding an athlete's calibration. Additionally, there are factors (e.g. gender) whose relationships with calibration might not always be visible without accounting for variance from other factors (e.g. performance), and factors whose influence weakens or disappears after accounting for the variance of other factors (e.g. experience). It is important to be aware of such cases, as they could determine whether we are actually using the most appropriate factors to assess athlete calibration. In conclusion, the present findings illustrate the utility of demographic factors in assessing exercise calibration, whilst they also highlight the need to exercise caution when doing so.

Prediction lead findings in Study 2 of Chapter 2 demonstrated that athletes are more likely to be precise in their predictions when they assess their performance closer to the time of a competition than earlier in advance. To my knowledge, there has been no previous research on the relationship between prediction lead and calibration, as previous studies either did not have information regarding the time when participants made their predictions (Hubble & Zhao, 2016; Liverakos et al., 2018), or collected predictions right before the event examined (Krawczyk & Wilamowski, 2016, 2018). These results thus provide initial evidence for the importance of accounting for the time when athletes make their performance predictions, as it appears to influence their prediction accuracy. This has substantial implications for training and competition strategies, as strategies developed earlier on are likely to depend on less precise performance estimates than strategies developed later in a training programme or closer to the time of a race. Therefore, athletes and coaches need to factor the contribution of prediction lead to calibration, and update training and competition programmes and strategies in line with the most recent performance estimates. Furthermore, event organisers should be aware of prediction lead when they ask athletes for their performance predictions, e.g. to determine starting placement, in order to receive estimates that will better reflect the performance capacities of athletes on the day of the event.

The studies in Chapter 2 were also, to my knowledge, the first to examine prediction type effects on running calibration, as previous studies typically collected data for only one prediction type (e.g. Krawczyk & Wilamowski, 2016, 2018). Interestingly, both studies in Chapter 2 reported that participants were underconfident in their realistic predictions, contrasting patterns of overconfidence previously observed in running calibration studies (Krawczyk & Wilamowski, 2016, 2018). Asking participants to provide both prediction types could have affected how they made their realistic and goal predictions, thus leading to this discrepancy in realistic predictions with previous research. Furthermore, though Study 2 exhibited overconfidence in goal predictions, Study 1 did not, which was an inconsistent finding with previous evidence of goal prediction overconfidence (Sackett et al., 2015). However, the lack of goal prediction overconfidence in Study 1 could have been situation-specific, as the sample exhibited a higher tendency towards underconfidence than Study 2 (see Section 2.4 for a possible explanation). These results highlight the importance of accounting for prediction type when we assess exercise calibration, as participants might produce different prediction types based on how they perceive prediction instructions (i.e. do the instructions ask for "goal" or "realistic" predictions?). It is thus important for researchers, coaches, and event organisers to be specific in how they want athletes to assess their performance to avoid prediction type confounds in calibration. For example, training and competition programmes and strategies, which assume that athletes make realistic estimates of their performance, might not be as effective when athletes opt to make goal predictions instead, leading to suboptimal performance outcomes. Similarly, event organisers might place athletes who make goal, rather than realistic, predictions in starting positions that do not necessarily reflect their performance capacity.

Results from Chapter 4 were important in understanding whether we can utilise self-reports of exercise metacognition and cognitive calibration to predict athlete calibration. Studies 4 and 5 found no association between self-reports of exercise metacognition and exercise calibration, contrasting the cognitive and exercise studies that have found individuals with higher scores on metacognition self-reports to exhibit a tendency to be better calibrated than individuals with lower scores (Jang et al., 2020; Nietfeld, 2003; Schraw, 1997; Tobias et al., 1999). However, they were in accordance with the studies that have observed a lack of association between offline measures of metacognition and calibration (Gutierrez & Schraw, 2015; Jacobse & Harskamp, 2012; Saraç & Karakelle, 2012; Schraw & Dennison, 1994; Sperling et al., 2004; Zepeda et al., 2015). This lack of association was not necessarily surprising in Study 4, as the self-report questionnaire used targeted general exercise metacognition, whilst participants did not have previous experience with the modality examined—they might have thus had limited relevant metacognitive knowledge and experience. It has been suggested that self-reports of metacognition need to be specific to the task used to measure calibration for a relationship with calibration to arise (Schellings, 2011; Schellings et al., 2013). The absence of significant findings in Study 5 was thus more surprising than it was in Study 4, as the self-report questionnaire used was specific to running and participants had previous running experience. Consequently, these results suggest that simply adjusting items in metacognitive questionnaires to refer to a specific exercise modality is likely an insufficient method to produce questionnaires that are specific enough to predict exercise calibration (Jacobse & Harskamp, 2012).

Study 6 also produced results that were not in line with the relevant—limited—literature. Though Arbuzova and colleagues (2020) found positive correlations in metacognitive efficiency between motor and cognitive conditions, this was not the case in Study 6. Bias and absolute accuracy in the memory recognition task did not correlate with bias and absolute accuracy in either HIFME workout. In fact, cognitive performance and performance judgments did not correlate with HIFME performance and predictions either. Though it is not clear why Study 6 showed a dissociation between cognitive and exercise calibration when Arbuzova and colleagues (2020) found a positive association, it is likely that methodological differences in the metacognitive measures and tasks used in the two studies contributed to this discrepancy. Contrasting Study 6, Study 4 exhibited a positive correlation between academic and exercise metacognition self-reports, which supported previous findings on the generalisability of metacognition self-reports across exercise and academic domains (Jonker et al., 2010, 2011; Mccardle, 2015). Thus, results from Studies 4 and 6 indicate that we are more likely to observe domain-generality across cognitive and exercise modalities in metacognition self-reports than in calibration measures.

Taken together, results from studies 4, 5, and 6 and the relevant literature suggest that the use of metacognition self-reports and cognitive calibration is not effective in predicting athlete calibration. This appears to be the case even when metacognition self-reports are specific to the calibration activity. The dissociation between metacognition self-reports and calibration is likely the result of methodological differences between the two measurement methods, e.g. participants can self-report metacognitive behaviour that does not reflect their actual behaviour when they predict their prospective performance (Harrison & Vallin, 2018; Tobias et al., 1999). One explanation for the lack of association between exercise and cognitive calibration in Study 6 could be that metacognitive processes do not generalise across the two domains. However, studies examining exercise metacognition, and investigating metacognition across different cognitive (and motor) tasks (e.g. Arbuzova et al., 2020; Brick et al., 2015; Carpenter et al., 2019; MacIntyre et al., 2014; Mazancieux et al., 2020; Rouault et al., 2018), point towards the existence of several domain-general metacognitive processes (e.g. performance monitoring and regulation), which we would expect to be consistent across domains. At the same time, research on the relationship between exercise and cognitive calibration is too scarce to draw reliable conclusions regarding theoretical differences in metacognitive processes between the two domains. A possible explanation for the results in Study 6 would be that the exercise and cognitive tasks used were very different from each other, and thus required the implementation of different metacognitive processes, leading to the observed calibration dissociation. Because of this possibility, as well as other methodological considerations (see Section 4.5), we need to further test the relationship between exercise and cognitive calibration using a wide range of tasks from each domain, as well as consistent metacognitive judgments and measures. Overall, though we need to conduct further research to gain a more thorough understanding of the relationships examined in Chapter 4 (see Section 6.4 for suggestions), data from the present thesis do not support the use of metacognition self-reports and cognitive calibration to assess exercise calibration. In contrast, findings from Study 4 support the use of scores from self-reports of academic metacognition to predict scores from self-reports of exercise metacognition, and vice versa, in recreational athletes.

Chapter 5 findings showed that even a minimal metacognitive manipulation in the form of prediction guidance can be effective in optimising prediction precision, regardless of whether athletes have previous experience with an exercise modality or not. This supports previous cognitive and exercise evidence on the positive effects of metacognitive interventions on calibration (e.g. Gutierrez & Schraw, 2015; Gutierrez de Blume, 2017; Kolovelonis et al., 2020; Nietfeld et al., 2006). Contrasting precision results, strategic instructions in Study 7, which asked participants to adjust for fatigue in their predictions, led to strategic prediction underconfidence, whilst impulsive predictions were unbiased. However, this was likely the result of the strategy described in the instructions being suboptimal and inappropriate for the workout examined. This became evident in Study 8, where minimal strategic instructions led to a lack of prediction bias and a lack of difference in bias with the impulsive condition. In line with these findings, physical education research has produced inconsistent results on the effects of self-regulation training on prediction bias (Kolovelonis, Goudas, & Dermitzaki, 2012; Kolovelonis et al., 2013, 2020). Taken together, present and previous results suggest that bias is sensitive to the specific metacognitive manipulations used (e.g. the specific instructions participants receive). Overall, Chapter 5 has provided us with important information on how we can optimise exercise calibration in a time-efficient manner, as well as on the necessity of selecting appropriate manipulations to achieve a positive calibration outcome.

6.4 FUTURE DIRECTIONS

The results I obtained in the present thesis have important implications for exercise calibration research. Until now, the literature on the ways in which we can assess and optimise athlete calibration has been limited. To address this, the studies I presented in Chapters 2, 3, 4, and 5 aimed to initiate a sustained effort to produce a comprehensive and informative framework of calibration in the domain of exercise. I discuss below some of the directions that future research can take based on the present thesis.

The associations between demographic factors and calibration were not consistent across running and HIFME in Chapters 2 and 3. Though we can partly attribute this to the study in Chapter 3 having a smaller sample size, and thus lower statistical power, than the studies in Chapter 2, it is also possible that the same demographic factors have a different predictive capacity of calibration across exercise modalities. To test this possibility, it is essential to conduct further research that explores the reliability of findings from one modality (e.g. running) to another (e.g. HIFME). This way we can assess whether using the same demographic factors to predict calibration across different exercise modalities is an effective approach to assessing athlete metacognition. Additionally, we should examine the reliability of these relationships within each exercise modality (e.g. by collecting data from running races of different lengths) in the event that different intra-activity demands and tasks alter associations between demographic factors and calibration.

It is also important to conduct further investigations on the relationships between demographic factors and exercise calibration, as there are factors and aspects of these relationships, which I did not examine in the present thesis. For example, it would be interesting to explore how and whether different markers of expertise (e.g. performance, competence rankings, and achievements in competitions) relate to each other, and test the extent to which certain types of expertise operationalisation are better at informing us about calibration than others. This way we could learn whether some markers of expertise are more useful than other when it comes to assessing an athlete’s exercise calibration. In the same vein, since experience is a multifaceted factor, we should explore the possibility that markers of experience other than the ones examined in Chapters 2 and 3 also have the capacity to predict calibration. In running, such factors could consist of the number of previous races (same or different race and course) participants have completed in the past; previous emphasis on competitive or recreational running; types of training (e.g. interval versus long-distance running); and whether participants have previous experience of engaging in metacognitive behaviour in running. Researching the associations of these factors with exercise calibration would better inform us on the role of experience in metacognition, and could assist us in inferring which markers of experience are the most effective in predicting athlete calibration. Based on this information, we could then devise experience questionnaires aiming to assist coaches, fitness instructors, and athletes in assessing athlete calibration tendencies.

To further our understanding of how age contributes to exercise calibration, it is important to test the reliability of findings in Chapter 2. To do so, we need to conduct more examinations on the relationships between age and exercise calibration, whilst ensuring that the samples we collect contain data from athletes across a wide range of ages. In terms of gender research in calibration, it is essential to examine and determine the best methods of accounting for gender differences in athletic performance capacity across exercise modalities. Simply accounting for performance variance that has not been adjusted for gender differences in performance capacity may lead to the overestimation of gender differences in bias (Deaner et al., 2014). There have thus been previous attempts to address this potential issue in running (i.e. by adjusting female finish time by a theoretical value of 12%; Deaner et al., 2014). However, before we can decide on what the optimal method of accounting for gender differences in performance capacity is, we need to first test and compare the validity and effectiveness of different methods, as well as examine whether we could generalise them across exercise modalities or whether we should develop modality-specific methods.

The studies in Chapter 2 were the first to explore the effects of prediction type on calibration. However, there were certain limitations that did not allow me to reach strong conclusions regarding this relationship. For example, since I did not use counterbalancing in the two studies, the order in which participants made their predictions could have biased the calibration results. This could have happened despite my instructions to consider both prediction types before making predictions. Thus, we need to conduct more research that counterbalances the order of prediction instructions in its examination of the effects of prediction type on exercise calibration. Furthermore, it would be interesting to explore the possibility of whether asking participants to provide both prediction types could lead to differences in calibration compared to asking them to provide only one prediction type. Such examinations will provide us with a better and clearer understanding of the role that prediction type plays in exercise calibration.

Findings from Chapter 4 illustrated that metacognition self-reports are ineffective in assessing exercise calibration, even if we use questionnaires that are specific to the exercise modality we use to assess calibration. However, it is possible that the running metacognition questionnaire I used in Study 5 was still not specific enough to inform us about running calibration, as I simply adjusted general exercise items to refer to running (Jacobse & Harskamp, 2012). Perhaps developing a metacognitive questionnaire with a specific exercise modality in mind, and ensuring that all items refer to calibration processes will lead to a different result, where metacognitive scores successfully predict calibration within the same modality. Given the potential benefits of such a relationship, it is worth testing the above possibility, even if it means that we need to develop specific metacognitive questionnaires for different exercise modalities. Along the same lines, despite cognitive calibration not correlating with HIFME calibration in Study 6, it is imperative that we still conduct more investigations of this relationship, where we use a variety of athletic, motor, and cognitive tasks, as well as a variety of metacognitive measures and judgments. Perhaps cognitive and athletic tasks that are more similar with each other compared to the tasks used in the present study will be more likely to exhibit calibration associations. Such research would also allow us to better understand the extent to which metacognitive processes generalise across the exercise and cognitive domains (for a more detailed discussion, see Section 6.3). Overall, we should use findings from Chapter 4 to inform and guide future research that examines the relationships of metacognition self-reports and cognitive calibration with exercise calibration, rather than to simply dismiss the possibility that these relationships exist altogether.

Finally, the studies in Chapter 5 were the first to provide evidence for the effects of prediction guidance on exercise calibration. Building on these findings, future research should aim to further explore this type of metacognitive intervention. For example, in the studies I presented, I did not examine how general and minimal instructions (such as the ones used in Study 8) impact calibration when participants lack previous experience with a modality, nor did I investigate the effect of specific strategic instructions (such as the ones used in Study 7) on calibration when participants have previous exercise experience. These effects are important to explore, as they will provide us with information on how we can best implement prediction guidance interventions to optimise exercise calibration. Additionally, using different exercise modalities of varying intensity, duration, and structure will inform us on the ways in which we can adjust prediction guidance instructions based on the activity for which athletes need to assess their performance. At the same time, it is important to investigate how athletes make their predictions when they do not receive instructions to be impulsive or strategic, and whether strategic instructions will still lead to better calibration compared to such an experimental condition. Some participants likely tend to engage in strategic thinking in their performance judgments without needing to receive specific instructions to do so. In such cases, it would be best for coaches and fitness instructors to focus on providing strategic instructions to athletes who have a tendency to make impulsive predictions, instead of focusing on athletes who are already likely to make strategic predictions.

6.5 CONCLUSION

The present thesis investigated and provided novel insights into the different ways in which we can predict and optimise athlete calibration in running and HIFME. In doing so, I found evidence on how a variety of demographic factors can inform us about an athlete's calibration. In contrast, I found no evidence to suggest that exercise metacognition self-reports and cognitive calibration play a similar role in exercise calibration. Additionally, I collected data that highlighted the importance of accounting for prediction type when asking participants to make their predictions. In the same vein, providing participants with strategic guidance on how to make their performance predictions appears to elicit a positive effect on exercise calibration compared to instructing them to be impulsive and non-strategic. These findings have important practical and theoretical implications for exercise calibration. In terms of practical implications, evidence points towards the use of demographic factors to inform instructors, event organisers, and athletes regarding athlete calibration tendencies. Based on this information, athletes can then engage in appropriate metacognitive behaviour to address possible patterns of miscalibration. Furthermore, when athletes do not engage in metacognitive behaviour to assess their prospective performance, instructors should provide them with appropriate strategic prediction guidance, aiming to optimise their calibration. In terms of theoretical implications, the present findings have made important contributions to expanding our understanding of exercise calibration, and should serve as a guide for future research in the field.

CHAPTER 7: REFERENCES

Albinet, C. T., Abou-Dest, A., André, N., & Audiffren, M. (2016). Executive functions improvement following a 5-month aquaerobics program in older adults: Role of cardiac vagal control in inhibition performance. *Biological Psychology*, *115*, 69–77. https://doi.org/10.1016/j.biopsycho.2016.01.010

Albinet, C. T., Boucard, G., Bouquet, C. A., & Audiffren, M. (2010). Increased heart rate variability and executive performance after aerobic training in the elderly. *European Journal of Applied Physiology*, *109*(4), 617–624. https://doi.org/10.1007/s00421-010-1393-y

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, *24*(1), 1–3. https://doi.org/10.1016/j.learninstruc.2012.10.003

Alvariñas-Villaverde, M., López-Villar, C., Fernández-Villarino, M. A., & Alvarez-Esteban, R. (2017). Masculine, feminine and neutral sports: Extracurricular sport modalities in practice. *Journal of Human Sport and Exercise*, *12*(4), 1278–1288. https://doi.org/10.14198/jhse.2017.124.14

Andersen, J. J. (2020). The State of Running 2019. Retrieved August 4, 2020, from https://runrepeat.com/state-of-running

Anderson-Hanley, C., Arciero, P. J., Brickman, A. M., Nimon, J. P., Okuma, N., Westen, S. C., … Zimmerman, E. A. (2012). Exergaming and older adult cognition: A cluster randomized clinical trial. *American Journal of Preventive Medicine*, *42*(2), 109–119. https://doi.org/10.1016/j.amepre.2011.10.016

Anderson, L., Oldridge, N., Thompson, D. R., Zwisler, A. D., Rees, K., Martin, N., & Taylor, R. S. (2016). Exercise-Based Cardiac Rehabilitation for Coronary Heart Disease Cochrane Systematic Review and Meta-Analysis. *Journal of the American College of Cardiology*, *67*(1), 1–12. https://doi.org/10.1016/j.jacc.2015.10.044

Arbuzova, P., Peters, C., Roed, L., Koss, C., Maurer, H., Maurer, L. K., … Filevich, E. (2020). Measuring Metacognition of Direct and Indirect Parameters of Voluntary Movement. *BioRxiv*, 2020.05.14.092189. https://doi.org/10.1101/2020.05.14.092189

Avila, J. A. de, Avila, R. A. de, Gonçalves, E. M., & Guerra-Junior, G. (2019). Influence of physical training on bone mineral density in healthy young adults: A systematic review. *Revista Da Associação Médica Brasileira*, *65*(8), 1102–1106. https://doi.org/http://dx.doi.org/10.1590/1806-9282.65.8.1102

Avugos, S., Bar-Eli, M., Ritov, I., & Sher, E. (2013). The elusive reality of efficacy–performance cycles in basketball shooting: An analysis of players' performance under invariant conditions. *International Journal of Sport and Exercise Psychology*, *11*(2), 184–202. https://doi.org/10.1080/1612197X.2013.773661

Baker, L. D., Frank, L. L., Foster-schubert, K., Green, P. S., Wilkinson, C. W., Mctiernan, A., … Craft, S. (2010). Effects of Aerobic Exercise on Mild Cognitive Impairment. *Arch Neurol.*, *67*(1), 71–79.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459.

Bashore, T. R., Ally, B., van Wouwe, N. C., Neimat, J. S., van den Wildenberg, W. P. M., & Wylie, S. A. (2018). Exposing an "intangible" cognitive skill among collegiate football players: II. enhanced response impulse control. *Frontiers in Psychology*, *9*, 1496. https://doi.org/10.3389/fpsyg.2018.01496

Boekaerts, M. (1996). Self-regulated Learning at the Junction of Cognition and Motivation. *European Psychologist*, *1*(2), 100–112.

Borg, G. A. V. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise*, *14*(5), 377–381. https://doi.org/10.1249/00005768-198205000-00012

Boucard, G. K., Albinet, C. T., Bugaiska, A., Bouquet, C. A., Clarys, D., & Audiffren, M. (2012). Impact of physical activity on executive functions in aging: A selective effect on inhibition among old adults. *Journal of Sport and Exercise Psychology*, *34*(6), 808–827. https://doi.org/10.1123/jsep.34.6.808

Breen, D., Norris, M., Healy, R., & Anderson, R. (2018). Marathon pace control in masters athletes. *International Journal of Sports Physiology and Performance*, *13*(3), 332–338. https://doi.org/https://doi.org/10.1123/ijspp.2016-0730

Brick, N. E., Campbell, M. J., Metcalfe, R. S., Mair, J. L., & Macintyre, T. E. (2016). Altering pace control and pace regulation: Attentional focus effects during running. *Medicine and Science in Sports and Exercise*, *48*(5), 879–886. https://doi.org/10.1249/MSS.0000000000000843

Brick, N. E., MacIntyre, T. E., & Campbell, M. J. (2016). Thinking and action: A cognitive perspective on self-regulation during endurance performance. *Frontiers in Physiology*, *7*, 159. https://doi.org/10.3389/fphys.2016.00159

Brick, N., MacIntyre, T., & Campbell, M. (2014). Attentional focus in endurance activity: new paradigms and future directions. *International Review of Sport and Exercise Psychology*, *7*(1), 106–134. https://doi.org/10.1080/1750984X.2014.885554

Brick, N., MacIntyre, T., & Campbell, M. (2015). Metacognitive processes in the self-regulation of performance in elite endurance runners. *Psychology of Sport and Exercise*, *19*, 1–9. https://doi.org/10.1016/j.psychsport.2015.02.003

Brown, A. K., Liu-Ambrose, T., Tate, R., & Lord, S. R. (2009). The effect of group-based exercise on cognitive performance and mood in seniors residing in intermediate care and self-care retirement facilities: a randomised controlled trial. *British Journal of Sports Medicine*, *43*(8), 608–614. https://doi.org/10.1136/bjsm.2008.049882

Brown, A. L., Smiley, S. S., & Lawton, S. Q. S. (1978). The effects of experience on the selection of suitable retrieval cues for studying texts. *Child Development*, *49*(3), 829–835.

Brown, M. Y., Graham, J. D., Innes, K. I., Harris, S., Flemington, A., & Bray, S. R. (2019). Effects of Prior Cognitive Exertion on Physical Performance: A Systematic Review and Meta-analysis. *Sports Medicine*, *50*(3), 497–529. https://doi.org/10.1007/s40279-019-01204-8

Butcher, S. J., Neyedly, T. J., Horvey, K. J., & Benko, C. R. (2015). Do physiological measures predict selected crossFit ® benchmark performance? *Open Access Journal of Sports Medicine*, *6*, 241–247. https://doi.org/10.2147/OAJSM.S88265

Carlsson, T., Assarsson, H., & Carlsson, M. (2016). The influence of sex, age, and race experience on pacing profiles during the 90 km Vasaloppet ski race. *Open Access Journal of Sports Medicine*, *7*, 11–19. https://doi.org/10.2147/oajsm.s101995

Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General*, *148*(1), 51–64. https://doi.org/10.1037/xge0000505

Carretti, B., Borella, E., Zavagnin, M., & Beni, R. De. (2011). Impact of metacognition and motivation on the efficacy of strategic memory training in older adults: Analysis of specific, transfer and maintenance effects. *Archives of Gerontology and Geriatrics*, *52*(3), e192–e197. https://doi.org/10.1016/j.archger.2010.11.004

Cassilhas, R. C., Viana, V. A. R., Grassmann, V., Santos, R. T., Santos, R. F., Tufik, S., & Mello, M. T. (2007). The impact of resistance exercise on the cognitive function of the elderly. *Medicine and Science in Sports and Exercise*, *39*(8), 1401–1407. https://doi.org/10.1249/mss.0b013e318060111f

Castel, A. D., Middlebrooks, C. D., & McGillivray, S. (2016). Monitoring Memory in Old Age: Impaired, Spared, and Aware. In J. Dunlosky & S. K. Tauber (Eds.), *Oxford Handbook of Metamemory* (pp. 519–534). New York, NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.3

Cauvin, S., Moulin, C. J. A., Souchay, C., Kliegel, M., & Schnitzspahn, K. M. (2019). Prospective Memory Predictions in Aging: Increased Overconfidence in Older Adults. *Experimental Aging Research*, *45*(5), 436–459. https://doi.org/10.1080/0361073x.2019.1664471

Chatzipanteli, A., & Digelidis, N. (2011). The Influence of Metacognitive Prompting on Students' Performance in a Motor Skills Test in Physical Education. *International Journal of Sports Science and Engineering*, *5*(2), 93–98.

Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, *20*(1), 2–17. https://doi.org/10.1016/j.learninstruc.2008.11.002

Claudino, J. G., Gabbett, T. J., Bourgeois, F., Souza, H. D. S., Miranda, R. C., Mezêncio, B., … Serrão, J. C. (2018). CrossFit Overview: Systematic Review and Meta-analysis. *Sports Medicine - Open*, *4*(11), 1–14.

Cleary, T., Zimmerman, B., & Keating, T. (2006). Training physical education students to self-regulate during basketball free-throw practice. *Research Quarterly for Exercise and Sport*, *77*(2), 251–262.

Colcombe, S. J., & Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults: a meta-analytic study. *Psychological Science*, *14*(2), 125–130. https://doi.org/10.1111/1467-9280.t01-1-01430

Coyle, E. F. (2007). Physiological Regulation of Marathon Performance. *Sports Medicine*, *37*(4–5), 306–311. https://doi.org/0112-1642/07/0004-0306/$44.95/0

Cuk, I., Nikolaidis, P. T., & Knechtle, B. (2020). Sex differences in pacing during half-marathon and marathon race. *Research in Sports Medicine*, *28*(1), 111–120. https://doi.org/10.1080/15438627.2019.1593835

Dahlbom, L., Jakobsson, A., Jakobsson, N., & Kotsadam, A. (2011). Gender and overconfidence: Are girls really overconfident? *Applied Economics Letters*, *18*(4), 325–327. https://doi.org/10.1080/13504851003670668

Dalgas, U., Stenager, E., Jakobsen, J., Petersen, T., Hansen, H., Knudsen, C., … Ingemann-Hansen, T. (2010). Fatigue, mood and quality of life improve in MS patients after progressive resistance training. *Multiple Sclerosis Journal*, *16*(4), 480–490. https://doi.org/10.1177/1352458509360040

de Assis, G. G., & de Almondes, K. M. (2017). Exercise-dependent BDNF as a modulatory factor for the executive processing of individuals in course of cognitive decline. A systematic review. *Frontiers in Psychology*, *8*, 1–8. https://doi.org/10.3389/fpsyg.2017.00584 de Bruin, A. B. H., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, *22*(4), 245–252. https://doi.org/10.1016/j.learninstruc.2012.01.003

Deaner, R. O., Addona, V., Carter, R. E., Joyner, M. J., & Hunter, S. K. (2016). Fast men slow more than fast women in a 10 kilometer road race. *PeerJ*, *4*, e2235. https://doi.org/10.7717/peerj.2235

Deaner, R. O., Addona, V., & Hanley, B. (2019). Risk taking runners slow more in the marathon. *Frontiers in Psychology*, *10*, 333. https://doi.org/10.3389/fpsyg.2019.00333

Deaner, R. O., Carter, R. E., Joyner, M. J., & Hunter, S. K. (2014). Men are more likely than women to slow in the marathon. *Medicine and Science in Sports and Exercise*, *47*(3), 607–616. https://doi.org/10.1249/MSS.0000000000000432

Deaner, R. O., & Lowen, A. (2016). Males and females pace differently in high school cross-country races. *Journal of Strength and Conditioning Research*, *30*(11), 2991–2997. https://doi.org/https://doi.org/10.1519/JSC.0000000000001407

Deci, E., & Ryan, R. (2008). Self-determination theory: A macrotheory of motivation, human development, and health. *Canadian Psychology*, *49*(3), 182–185. https://doi.org/https://doi.org/10.1037/a0012801

del Coso, J., Salinero, J. J., Lara, B., Abian-Vicen, J., Gallo-Salazar, C., & Areces, F. (2017). A comparison of the physiological demands imposed by competing in a half-marathon vs. a marathon. *The Journals of Sports Medicine and Physical Fitness*, *57*(11), 1399–1406. https://doi.org/10.23736/S0022-4707.17.07056-6

Dempsey, J. A., Amann, M., Romer, L. M., & Miller, J. D. (2008). Respiratory system determinants of peripheral fatigue and endurance performance. *Medicine and Science in Sports and Exercise*, *40*(3), 457–461. https://doi.org/10.1249/MSS.0b013e31815f8957

Devolder, P. A., Brigham, M. C., & Pressley, M. (1990). Memory performance awareness in younger and older adults. *Psychology and Aging*, *5*(2), 291–303. https://doi.org/10.1037/0882-7974.5.2.291

Díaz, J. J., Fernández-Ozcorta, E. J., & Santos-Concejero, J. (2018). The influence of pacing strategy on marathon world records. *European Journal of Sport Science*, *18*(6), 781–786. https://doi.org/10.1080/17461391.2018.1450899

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging*, *22*(1), 122–133. https://doi.org/10.1037/0882-7974.22.1.122

Drew, M. K., Finch, C. F., & Drew, M. K. (2016). The Relationship Between Training Load and Injury, Illness and Soreness: A Systematic and Literature Review. *Sports Medicine*, *46*(6), 861–883. https://doi.org/10.1007/s40279-015-0459-8

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognise their own incompetence. *Current Directions in Psychological Science: A Journal of the American Psychological Society*, *12*, 83. https://doi.org/10.1111/1467-8721.01235

Eckard, T. G., Padua, D. A., Hearn, D. W., Pexa, B. S., & Frank, B. S. (2018). The Relationship Between Training Load and Injury in Athletes: A Systematic Review. *Sports Medicine*, *48*(8), 1929–1961. https://doi.org/10.1007/s40279-018-0951-z

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist*, *13*(4), 277–287. https://doi.org/10.1027/1016-9040.13.4.277

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, *46*(1), 6–25. https://doi.org/10.1080/00461520.2011.538645

Efklides, A. (2014). How Does Metacognition Contribute to the Regulation of Learning? An Integrative Approach. *Psychological Topics*, *23*(1), 1–30.

Elferink-Gemser, M. T., & Hettinga, F. J. (2017). Pacing and self-regulation: Important skills for talent development in endurance sports. *International Journal of Sports Physiology and Performance*, *12*(6), 831–835. https://doi.org/10.1123/ijspp.2017-0080

Enoka, R. M., & Stuart, D. G. (1992). Neurobiology of muscle fatigue. *Journal of Applied Physiology*, *72*(5), 1631–1648. https://doi.org/10.1152/jappl.1992.72.5.1631

Erickson, K. I., Colcombe, S. J., Elavsky, S., McAuley, E., Korol, D. L., Scalf, P. E., & Kramer, A. F. (2007). Interactive effects of fitness and hormone treatment on brain health in postmenopausal women. *Neurobiology of Aging*, *28*(2), 179–185. https://doi.org/10.1016/j.neurobiolaging.2005.11.016

Fiuza-Luces, C., Santos-Lozano, A., Joyner, M., Carrera-Bastos, P., Picazo, O., Zugaza, J. L., … Lucia, A. (2018). Exercise benefits in cardiovascular disease: beyond attenuation of traditional risk factors. *Nature Reviews Cardiology*, *15*(12), 731–743. https://doi.org/10.1038/s41569-018-0065-1

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231–235). Hillsdale, NJ: Lawrence Erlbaum.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 1–9. https://doi.org/10.3389/fnhum.2014.00443

Fogarty, G. J., & Else, D. (2005). Performance calibration in sport: Implications for self-confidence and metacognitive biases. *International Journal of Sport and Exercise Psychology*, *3*(1), 41–57. https://doi.org/10.1080/1612197X.2005.9671757

Fogarty, G. J., & Ross, A. (2007). Calibration in tennis: The role of feedback and expertise. In K. Moore (Ed.), *Proceedings of the 42nd Australian Psychological Society Annual Conference: Psychology Making an Impact* (pp. 148–152). Brisbane, Australia.

Foster, G. D., Wadden, T. A., Vogt, R. A., & Brewer, G. (1997). What is a reasonable weight loss? Patients' expectations and evaluations of obesity treatment outcomes. *Journal of Consulting and Clinical Psychology*, *65*(1), 79–85. https://doi.org/10.1037/0022-006X.65.1.79

Franklin, B. A., Forgac, T. M., & Hellerstein, H. K. (1978). Accuracy of predicted marathon time: Relationship of training mileage to performance. *Research Quarterly of the American Alliance for Health, Physical Education and Recreation*, *49*(4), 450–459. https://doi.org/10.1080/10671315.1978.10615559

Gabbett, T. J., & Domrow, N. (2007). Relationships between training load, injury, and fitness in sub-elite collision sport athletes collision sport athletes. *Journal of Sports Sciences*, *25*(13), 1507–1519. https://doi.org/10.1080/02640410701215066

Gabbett, T. J., & Jenkins, D. G. (2011). Relationship between training load and injury in professional rugby league players. *Journal of Science and Medicine in Sport*, *14*(3), 204–209. https://doi.org/10.1016/j.jsams.2010.12.002

Gabr, S. A. (2019). Comments: "Influence of physical training on bone mineral density in healthy young adults: A systematic review." *Revista Da Associacao Medica Brasileira*, *65*(8), 1107–1108. https://doi.org/10.1590/1806-9282.65.8.1107

Gianzina, E. A., & Kassotaki, O. A. (2019). The benefits and risks of the high-intensity CrossFit training. *Sport Sciences for Health*, *15*(1), 21–33. https://doi.org/10.1007/s11332-018-0521-7

Gonida, E. N., & Leondari, A. (2011). Patterns of motivation among adolescents with biased and accurate self-efficacy beliefs. *International Journal of Educational Research*, *50*(4), 209–220. https://doi.org/10.1016/j.ijer.2011.08.002

Gosztyla, A. E., Edwards, D. G., Quinn, T. J., & Kenefick, R. W. (2006). The impact of different pacing strategies on five-kilometer running time trial performance. *Journal of Strength and Conditioning Research*, *20*(4), 882–886. https://doi.org/10.1519/R-19275.1

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory and Cognition*, *36*(1), 93–103. https://doi.org/10.3758/MC.36.1.93

Gutierrez, A. P., & Price, A. F. (2017). Calibration Between Undergraduate Students' Prediction of and Actual Performance: The Role of Gender and Performance Attributions. *Journal of Experimental Education*, *85*(3), 486–500. https://doi.org/10.1080/00220973.2016.1180278

Gutierrez, A. P., & Schraw, G. (2015). Effects of Strategy Training and Incentives on Students' Performance, Confidence, and Calibration. *The Journal of Experimental Education*, *83*(3), 386–404. https://doi.org/10.1080/00220973.2014.907230

Gutierrez de Blume, A. P. (2017). The effects of strategy training and an extrinsic incentive on fourth- and fifth-grade students' performance, confidence, and calibration accuracy. *Cogent Education*, *4*(1), 1314652. https://doi.org/10.1080/2331186X.2017.1314652

Hacker, D. J., Bol, L., & Keener, M. C. (2012). Metacognition in Education: A Focus on Calibration. In J. Dunlosky & R. Bjork (Eds.), *Handbook of Memory and Metacognition* (pp. 429–455). Mahwah, NJ: Larwrence Erlbaum Associates.

Händel, M., de Bruin, A. B. H., & Dresel, M. (2020). Individual differences in local and global metacognitive judgments. *Metacognition and Learning*, *15*, 51–75. https://doi.org/10.1007/s11409-020-09220-0

Hansson, P., Rönnlund, M., Juslin, P., & Nilsson, L. G. (2008). Adult Age Differences in the Realism of Confidence Judgments: Overconfidence, Format Dependence, and Cognitive Predictors. *Psychology and Aging*, *23*(3), 531–544. https://doi.org/10.1037/a0012782

Harrison, G. M., & Vallin, L. M. (2018). Evaluating the metacognitive awareness inventory using empirical factor-structure evidence. *Metacognition and Learning*, *13*(1), 15–38. https://doi.org/10.1007/s11409-017-9176-z

Haskell, W. L., Lee, I., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. A., … Bauman, A. (2007). Physical Activity and Public Health: Updated Recommendation for Adults from the American College of Sports Medicine and the American Heart Association. *Medicine & Science in Sports & Exercise*, *39*(8), 1423–1434. https://doi.org/10.1249/mss.0b013e3180616b27

Hazell, T. J., Hamilton, C. D., Olver, T. D., & Lemon, P. W. R. (2014). Running sprint interval training induces fat loss in women. *Applied Physiology, Nutrition, and Metabolism*, *39*(8), 944–950. https://doi.org/10.1139/apnm-2013-0503

Henderson, S. (2019). So How Many People Participated in the 2019 CrossFit Open? Retrieved August 24, 2020, from https://morningchalkup.com/2019/04/01/so-how-many-people-participated-in-the-2019-crossfit-open/#

Henry, J. D., MacLeod, M. S., Phillips, L. H., & Crawford, J. R. (2004). A Meta-Analytic Review of Prospective Memory and Aging. *Psychology and Aging*, *19*(1), 27–39. https://doi.org/10.1037/0882-7974.19.1.27

Hertzog, C., & Dunlosky, J. (2011). Metacognition in Later Adulthood: Spared Monitoring Can Benefit Older Adults' Self-Regulation. *Current Directions in Psychological Science*, *20*(3), 167–173. https://doi.org/10.1177/0963721411409026

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and Monitoring Associative Learning: Is Monitoring Accuracy Spared or Impaired? *Psychology and Aging*, *17*(2), 209–225. https://doi.org/10.1037//0882-7974.17.2.209

Hertzog, C., Sinclair, S. M., & Dunlosky, J. (2010). Age Differences in the Monitoring of Learning: Cross-Sectional Evidence of Spared Resolution Across the Adult Life Span. *Developmental Psychology*, *46*(4), 939–948. https://doi.org/10.1037/a0019812

Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: exercise effects on brain and cognition. *Nature*, *9*, 58–65.

Horgan, D. D. (1992). Children and chess expertise: The role of calibration. *Psychological Research*, *54*(1), 44–50.

Hubble, C., & Zhao, J. (2016). Gender differences in marathon pacing and performance prediction. *Journal of Sports Analytics*, *2*(1), 19–36. https://doi.org/10.3233/jsa-150008

Huff, J. D., & Nietfeld, J. L. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition and Learning*, *4*(2), 161–176. https://doi.org/10.1007/s11409-009-9042-8

Jacobs, J. E., & Paris, S. G. (1987). Children's Metacognition About Reading: Issues in Definition, Measurement, and Instruction. *Educational Psychologist*, *22*(3–4), 255–278. https://doi.org/10.1080/00461520.1987.9653052

Jacobse, A. E., & Harskamp, E. G. (2012). Towards efficient measurement of metacognition in mathematical problem solving. *Metacognition and Learning*, *7*(2), 133–149. https://doi.org/10.1007/s11409-012-9088-x

Jakobsson, N. (2012). Gender and confidence: Are women underconfident? *Applied Economics Letters*, *19*(11), 1057–1059. https://doi.org/10.1080/13504851.2011.613745

Jakobsson, N., Levin, M., & Kotsadam, A. (2013). Gender and Overconfidence: Effects of Context, Gendered Stereotypes, and Peer Group. *Advances in Applied Sociology*, *3*(2), 137–141. https://doi.org/10.4236/aasoci.2013.32018

Jang, Y., Lee, H., Kim, Y., & Min, K. (2020). The Relationship between Metacognitive Ability and Metacognitive Accuracy. *Metacognition and Learning*, 1–24. https://doi.org/https://doi.org/10.1007/s11409-020-09232-w

Jeffery, R. W., Forster, J. L., French, S. A., Kelder, S. H., Lando, H. A., McGovern, P. G., … Baxter, J. E. (1993). The healthy worker project: A work-site intervention for weight control and smoking cessation. *American Journal of Public Health*, *83*(3), 395–401. https://doi.org/10.2105/AJPH.83.3.395

Jones, A. M., & Doust, J. H. (1996). A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. *Journal of Sports Sciences*, *14*(4), 321–327. https://doi.org/10.1080/02640419608727717

Jones, C. M., Griffiths, P. C., & Mellalieu, S. D. (2017). Training Load and Fatigue Marker Associations with Injury and Illness: A Systematic Review of Longitudinal Studies. *Sports Medicine*, *47*(5), 943–974. https://doi.org/10.1007/s40279-016-0619-5

Jonker, L., Elferink-Gemser, M. T., Toering, T. T., Lyons, J., & Visscher, C. (2010). Academic performance and self-regulatory skills in elite youth soccer players. *Journal of Sports Sciences*, *28*(14), 1605–1614. https://doi.org/https://doi.org/10.1080/02640414.2010.516270

Jonker, L., Elferink-Gemser, M. T., & Visscher, C. (2011). The role of self-regulatory skills in sport and academic performances of elite youth athletes. *Talent Development and Excellence*, *3*(2), 263–275.

Jonker, L., Gemser, M. T. E., & Visscher, C. (2009). Talented athletes and academic achievements: a comparison over 14 years. *High Ability Studies*, *20*(1), 55–64. https://doi.org/10.1080/13598130902863691

Karst, K., Dotzel, S., & Dickhäuser, O. (2018). Comparing global judgments and specific judgments of teachers about students' knowledge: Is the whole the sum of its parts? *Teaching and Teacher Education*, *76*, 194–203. https://doi.org/10.1016/j.tate.2018.01.013

Kelley, G. A., Kelley, K. S., & Pate, R. R. (2019). Exercise and adiposity in overweight and obese children and adolescents : a systematic review with network meta-analysis of randomised trials. *BMJ Open*, *9*, e031220. https://doi.org/10.1136/bmjopen-2019-031220

Kitsantas, A., Steen, S., & Huie, F. (2009). The role of self-regulated strategies and goal orientation in predicting achievement of elementary school children. *International Electronic Journal of Elementary Education*, *2*(1), 65–81.

Knechtle, B., Rüst, C. A., Rosemann, T., Knechtle, P., & Lepers, R. (2012). Age, Training, and Previous Experience Predict Race Performance in Long-Distance Inline Skaters, Not Anthropometry. *Perceptual and Motor Skills*, *114*(1), 141–156. https://doi.org/10.2466/05.PMS.114.1.141-156

Knechtle, B., Valeri, F., Zingg, M. A., Rosemann, T., & Rüst, C. A. (2014). What is the age for the fastest ultra-marathon performance in time-limited races from 6 h to 10 days? *Age*, *36*, 9715. https://doi.org/10.1007/s11357-014-9715-3

Kolovelonis, A. (2019). Relating students ' participation in sport out of school and performance calibration in physical education. *Issues in Educational Research*, *29*(3), 774–789.

Kolovelonis, A., & Goudas, M. (2013). The development of self-regulated learning of motor and sport skills in physical education: A review. *Hellenic Journal of Psychology*, *10*(3), 193–210.

Kolovelonis, A., & Goudas, M. (2018). The relation of physical self-perceptions of competence, goal orientation, and optimism with students' performance calibration in physical education. *Learning and Individual Differences*, *61*, 77–86. https://doi.org/10.1016/j.lindif.2017.11.013

Kolovelonis, A., Goudas, M., & Dermitzaki, I. (2012). Students' Performance Calibration in a Basketball Dibbling Task in Elementary Physical Education. *International Electronic Journal of Elementary Education*, *4*(3), 507–517.

Kolovelonis, A., Goudas, M., Dermitzaki, I., & Kitsantas, A. (2013). Self-regulated learning and performance calibration among elementary physical education students. *European Journal of Psychology of Education*, *28*(3), 685–701. https://doi.org/10.1007/s 10212-012-0135-4

Kolovelonis, A., Goudas, M., Hassandra, M., & Dermitzaki, I. (2012). Self-regulated learning in physical education: Examining the effects of emulative and self-control practice. *Psychology of Sport and Exercise*, *13*(4), 383–389. https://doi.org/10.1016/j.psychsport.2012.01.005

Kolovelonis, A., Goudas, M., & Samara, E. (2020). The Effects of a Self-Regulated Learning Teaching Unit on Students' Performance Calibration, Goal Attainment, and Attributions in Physical Education The Effects of a Self-Regulated Learning Teaching Unit on. *The Journal of Experimental Education*, 1–18. https://doi.org/10.1080/00220973.2020.1724852

Kramer, A. F., & Erickson, K. I. (2007). Capitalizing on cortical plasticity: influence of physical activity on cognition and brain function. *Trends in Cognitive Sciences*, *11*(8), 342–348. https://doi.org/10.1016/j.tics.2007.06.009

Kramer, A. F., Erickson, K. I., & Colcombe, S. J. (2006). Exercise, cognition, and the aging brain. *J Appl Physiol*, *101*, 1237–1242. https://doi.org/10.1152/japplphysiol.000500.2006.

Krätzig, G. P., & Arbuthnott, K. D. (2009). Metacognitive learning: The effect of item-specific experience and age on metamemory calibration and planning. *Metacognition and Learning*, *4*(2), 125–144. https://doi.org/10.1007/s11409-009-9039-3

Krawczyk, M., & Wilamowski, M. (2016). Are We All Overconfident in the Long Run? Evidence from One Million Marathon Participants. *Journal of Behavioral Decision Making*, *30*(3), 719–730. https://doi.org/10.1002/bdm.1984

Krawczyk, M., & Wilamowski, M. (2018). Task difficulty and overconfidence. Evidence from distance running. *Journal of Economic Psychology*, *75*(B), 102128. https://doi.org/10.1016/j.joep.2018.12.002

Kruger, J., & Dunning, D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lai, E. R. (2011). Motivation: A Literature Review Research. *Pearson Research Report*.

Landine, J., & Stewart, J. (1998). Relationship between metacognition, motivation, locus of control, self-efficacy, and academic achievement. *Canadian Journal of Counselling*, *32*(3), 200–212.

Larson, E. B., Wang, L., Bowen, J. D., Mccormick, W. C., Teri, L., & Crane, P. (2006). Exercise Is Associated with Reduced Risk for Incident Dementia among Persons 65 Years of Age and Older. *Annals of Internal Medicine*, *144*(2), 73–82. https://doi.org/https://doi.org/10.7326/0003-4819-144-2-200601170-00004

Latham, N. K., Bennett, D. A., Stretton, C. M., & Anderson, C. S. (2004). Systematic Review of Progressive Resistance Strength Training in Older Adults. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *59*(1), M48–M61. https://doi.org/10.1093/gerona/59.1.M48

Leckie, R. L., Oberlin, L. E., Voss, M. W., Prakash, R. S., Szabo-Reed, A., Chaddock-Heyman, L., … Erickson, K. I. (2014). BDNF mediates improvements in executive function following a 1-year exercise intervention. *Frontiers in Human Neuroscience*, *8*, 985. https://doi.org/10.3389/fnhum.2014.00985

Lepers, R., Piacentini, M. F., Reda, D., Minganti, C., Baldassare, R., Tarperi, C., … La Torre, A. (2019). Pacing profiles of master athletes according to their predicted finishing time. *Movement & Sport Sciences*, *2*(104), 37–44.

Lin, L. M., & Zabrucky, K. M. (1998). Calibration of Comprehension: Research and Implications for Education and Instruction. *Contemporary Educational Psychology*, *23*(4), 345–391. https://doi.org/10.1006/ceps.1998.0972

Lin, L., Zabrucky, K. M., & Moore, D. (2002). Effects of Text Difficulty and Adults' Age on Relative Calibration of Comprehension. *American Journal of Psychology*, *115*(2), 187–198. https://doi.org/10.2307/1423434

Lirgg, C. D. (1991). Gender Differences In Self-Confidence in Physical Activity: A Meta-Analysis of Recent Studies. *Journal of Sport and Exercise Psychology*, *13*(3), 294–310. https://doi.org/10.1123/jsep.13.3.294

Liu-Ambrose, T., Nagamatsu, L., Graf, P., Beattie, B., Ashe, M., & Handy, T. (2010). Resistance training and executive functions. *Archives of Internal Medicine*, *170*(2), 170–178. https://doi.org/10.1001/archinternmed.2009.494

Liverakos, K., McIntosh, K., Moulin, C. J. A., & O'Connor, A. R. (2018). How accurate are runners' prospective predictions of their race times? *PLoS ONE*, *13*(8), e0200744. https://doi.org/10.1371/journal.pone.0200744

Livingston, J. A. (2003). Metacognition: An Overview.

Lundeberg, M. A., Fox, P. W., & Punćochaŕ, J. (1994). Highly Confident but Wrong: Gender Differences and Similarities in Confidence Judgments. *Journal of Educational Psychology*, *86*(1), 114–121. https://doi.org/10.1037/0022-0663.86.1.114

MacIntyre, T. E., Igou, E. R., Campbell, M. J., Moran, A. P., & Matthews, J. (2014). Metacognition and action: A new pathway to understanding social and cognitive aspects of expertise in sport. *Frontiers in Psychology*, *5*, 1155. https://doi.org/10.3389/fpsyg.2014.01155

Mahmood, K. (2016). Do People Overestimate Their Information Literacy Skills? A Systematic Review of Empirical Evidence on the Dunning-Kruger Effect. *Communications in Information Literacy*, *10*(2), 199–213. https://doi.org/https://doi.org/10.15760/ comminfolit.2016.10.2.24

Malone, S., Hughes, B., Doran, D. A., Collins, K., & Gabbett, T. J. (2019). Can the workload–injury relationship be moderated by improved strength, speed and repeated-sprint qualities? *Journal of Science and Medicine in Sport*, *22*(1), 29–34. https://doi.org/10.1016/j.jsams.2018.01.010

March, D. S., Vanderburgh, P. M., TitleBaum, P. J., & Hoops, M. L. (2011). Age, sex, and finish time as determinants of pacing in the marathon. *Journal of Strength and Conditioning Research*, *25*(2), 386–391. https://doi.org/doi: 10.1519/JSC.0b013e3181bffd0f

Marcora, S. (2009). Perception of effort during exercise is independent of afferent feedback from skeletal muscles, heart, and lungs. *Journal of Applied Physiology*, *106*(6), 2060–2062. https://doi.org/10.1152/japplphysiol.90378.2008

Marcora, S. M. (2010). Counterpoint: Afferent feedback from fatigued locomotor muscles is not an important determinant of endurance exercise performance. *Journal of Applied Physiology*, *108*(2), 454–456. https://doi.org/10.11 52/japplphysiol.00976.2009a

Marcora, S. M., Staiano, W., & Manning, V. (2009). Mental fatigue impairs physical performance in humans. *Journal of Applied Physiology*, *106*(3), 857–864. https://doi.org/10.1152/japplphysiol.91324.2008

Marcora, Samuele Maria, & Staiano, W. (2010). The limit to exercise tolerance in humans: Mind over muscle? *European Journal of Applied Physiology*, *109*(4), 763–770. https://doi.org/10.1007/s00421-010-1418-6

Marzetti, E., Calvani, R., Tosato, M., Cesari, M., Di Bari, M., Cherubini, A., … Landi, F. (2017). Physical activity and exercise as countermeasures to physical frailty and sarcopenia. *Aging Clin Exp Res*, *29*, 35–42. https://doi.org/10.1007/s40520-016-0705-4

Mazancieux, A., Fleming, S. M., Souchay, C., & Moulin, C. J. A. (2020). Is There a G Factor for Metacognition? Correlations in Retrospective Metacognitive Sensitivity Across Tasks. *Journal of Experimental Psychology: General*. https://doi.org/http://dx.doi.org/10.1037/xge0000746

Mccardle, L. (2015). Similarities and Differences in Self-Regulated Learning Processes in Sport and Academics: A Case Study. *Journal for the Study of Sports and Athletes in Education*, *9*(3), 190–213. https://doi.org/10.1080/19357397.2015.1123002

McCormick, A., Meijen, C., Anstiss, P. A., & Jones, H. S. (2019). Self-regulation in endurance sports: theory, research, and practice. *International Review of Sport and Exercise Psychology*, *12*(1), 235–264. https://doi.org/https://doi.org/10.1002/bdm.472

Mcgillivray, S., & Castel, A. D. (2011). Betting on Memory Leads to Metacognitive Improvement by Younger and Older Adults. *Psychology and Aging*, *26*(1), 137–142. https://doi.org/10.1037/a0022681

McGraw, P. A., Mellers, B. A., & Ritov, I. (2004). The Affective Costs of Overconfidence. *Journal of Behavioral Decision Making*, *17*(4), 281–295. https://doi.org/https://doi.org/10.1002/bdm.472

Mellers, B. A. (2000). Choice and the Relative Pleasure of Consequences. *Psychological Bulletin*, *126*(6), 910–924. https://doi.org/https://doi.org/10.1037/0033-2909.126.6.910

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision Affect Theory: Emotional Reactions to the Outcomes of Risky Options. *Psychological Science*, *8*(6), 423–429. https://doi.org/https://doi.org/10.1111/j.1467-9280.1997.tb00455.x

Mellers, B., Schwartz, A., & Ritov, I. (1999). Emotion-Based Choice. *Journal of Experimental Psychology: General*, *128*(3), 332–345. https://doi.org/https://doi.org/10.1037/0096-3445.128.3.332

Meyer, J., Morrison, J., & Zuniga, J. (2017). The Benefits and Risks of CrossFit: A Systematic Review. *Workplace Health and Safety*, *65*(12), 612–618. https://doi.org/10.1177/2165079916685568

Model D. (n.d.). Retrieved August 24, 2020, from https://www.concept2.com/indoor-rowers/model-d

Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *The Journal of Neuroscience*, *38*(14), 3534–3546. https://doi.org/10.1523/jneurosci.2360-17.2018

Nagamatsu, L. S., Handy, T. C., Hsu, L. C., Voss, M., & Liu-Ambrose, T. (2012). Resistance training promotes cognitive and functional brain plasticity in seniors with probable mild cognitive impairment. *Archives of Internal Medicine*, *172*(8), 666–668. https://doi.org/10.1001/archinternmed.2012.379

Nederhand, M. L., Tabbers, H. K., Splinter, T. A. W., & Rikers, R. M. J. P. (2018). The Effect of Performance Standards and Medical Experience on Diagnostic Calibration Accuracy. *Health Professions Education*, *4*(4), 300–307. https://doi.org/10.1016/j.hpe.2017.12.008

Nekby, L., Thoursie, P. S., & Vahtrik, L. (2008). Gender and self-selection into a competitive environment: Are women more overconfident than men? *Economics Letters*, *100*(3), 405–407. https://doi.org/10.1016/j.econlet.2008.03.005

Nelson, T. O., & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. In G. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 26, pp. 125–173). New York: Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Nietfeld, J. (2003). An Examination of Metacognitive Strategy Use and Monitoring Skills by Competitive Middle Distance Runners. *Journal of Applied Sport Psychology*, *15*(4), 307–320. https://doi.org/10.1080/714044199

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning*, *1*(2), 159–179. https://doi.org/10.1007/s10409-006-9595-6

Nietfeld, J. L., & Schraw, G. (2002). The Effect of Knowledge and Strategy Training on Monitoring Accuracy. *The Journal of Educational Research*, *95*(3), 131–142. https://doi.org/10.1080/00220670209596583

Nikolaidis, P. T., Cuk, I., & Knechtle, B. (2019). Pacing of Women and Men in Half-Marathon and Marathon Races. *Medicina*, *55*(1), 14. https://doi.org/10.3390/medicina55010014

Nikolaidis, P. T., Cuk, I., Rosemann, T., & Knechtle, B. (2019). Performance and Pacing of Age Groups in Half-Marathon and Marathon. *International Journal of Environmental Research and Public Health*, *16*, 1777. https://doi.org/doi:10.3390/ijerph16101777

Nikolaidis, P. T., & Knechtle, B. (2017). Effect of age and performance on pacing of marathon runners. *Open Access Journal of Sports Medicine*, *8*, 171–180. https://doi.org/10.2147/oajsm.s141649

Nikolaidis, P. T., & Knechtle, B. (2018a). Pacing in age group marathoners in the "New York City Marathon." *Research in Sports Medicine*, *26*(1), 86–99. https://doi.org/10.1080/15438627.2017.1393752

Nikolaidis, P. T., & Knechtle, B. (2018b). Pacing Strategies in the 'Athens Classic Marathon': Physiological and Psychological Aspects. *Frontiers in Physiology*, *9*, 1539. https://doi.org/10.3389/fphys.2018.01539

Norman, E., Pfuhl, G., Sæle, R. G., Svartdal, F., Låg, T., & Dahl, T. I. (2019). Metacognition in Psychology. *Review of General Psychology*, *23*(4), 403–424. https://doi.org/10.1177/1089268019883821

Northey, J. M., Cherbuin, N., Pumpa, K. L., Smee, D. J., & Rattray, B. (2018). Exercise interventions for cognitive function in adults older than 50: A systematic review with meta-analysis. *British Journal of Sports Medicine*, *52*(3), 154–160. https://doi.org/10.1136/bjsports-2016-096587

Oettingen, G., & Wadden, T. A. (1991). Expectation, fantasy, and weight loss: Is the impact of positive thinking always positive? *Cognitive Therapy and Research*, *15*(2), 167–175. https://doi.org/10.1007/BF01173206

Okay, D. M., Jackson, P. V., Marcinkiewicz, M., & Papino, M. N. (2009). Exercise and Obesity. *Primary Care - Clinics in Office Practice*, *36*(2), 379–393. https://doi.org/10.1016/j.pop.2009.01.008

Ozsoy, G., Memis, A., & Temur, T. (2009). Metacognition, study habits and attitudes. *International Electronic Journal of Elementary Education*, *2*(1), 154–166.

Ozturk, N. (2017). Assessing Metacognition: Theory and Practices. *International Journal of Assessment Tools in Education*, *4*(2), 134–148. https://doi.org/10.21449/ijate.298299

Pageaux, B., & Lepers, R. (2018). The effects of mental fatigue on sport-related performance. In *Progress in Brain Research* (Vol. 240, pp. 291–315). Elsevier B.V. https://doi.org/10.1016/bs.pbr.2018.10.004

Pageaux, B., Lepers, R., Dietz, K. C., & Marcora, S. M. (2014). Response inhibition impairs subsequent self-paced endurance performance. *European Journal of Applied Physiology*, *114*(5), 1095–1105. https://doi.org/10.1007/s00421-014-2838-5

Paine, J., Uptgraft, J., & Wylie, R. (2010). CrossFit Study. *Command and General Staff College*, 1–34.

Pajares, F., & Graham, L. (1999). Self-efficacy, motivational constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, *24*(2), 124–139. https://doi.org/https://doi.org/10.1006/ceps.1998.0991

Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition*, *28*, 151–160. https://doi.org/10.1016/j.concog.2014.06.007

Paris, S. G., & Oka, E. R. (1986). Children’s reading strategies, metacognition, and motivation. *Developmental Review*, *6*(1), 25–56. https://doi.org/https://doi.org/10.1016/0273-2297(86)90002-X

Patel, H., Alkhawam, H., Madanieh, R., Shah, N., Kosmas, C. E., & Vittorio, T. J. (2017). Aerobic vs anaerobic exercise training effects on the cardiovascular system. *World Journal of Cardiology*, *9*(2), 134–138. https://doi.org/10.4330/wjc.v9.i2.134

Pedersen, B. K. (2017). Anti-inflammatory effects of exercise: role in diabetes and cardiovascular disease. *European Journal of Clinical Investigation*, *47*(8), 600–611. https://doi.org/10.1111/eci.12781

Peig-Chiello, P., Perrig, W. J., Ehrsam, R., Staehelin, H. B., & Krings, F. (1998). The effects of resistance training on well-being and memory in elderly volunteers. *Age and Ageing*, *27*(4), 469–475. https://doi.org/10.1093/ageing/27.4.469

Ramdass, D., & Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics*, *20*(1), 18–41. https://doi.org/10.4219/jaa-2008-869

Rapoport, B. I. (2010). Metabolic Factors Limiting Performance in Marathon Runners. *PLoS Computational Biology*, *6*(10), e1000960. https://doi.org/10.1371/journal.pcbi.1000960

Rogalski, B., Dawson, B., Heasman, J., & Gabbett, T. J. (2013). Training and game loads and injury risk in elite Australian footballers. *Journal of Science and Medicine in Sport*, *16*(6), 499–503. https://doi.org/10.1016/j.jsams.2012.12.004

Ross, R., Stotz, P. J., & Lam, M. (2015). Effects of exercise amount and intensity on abdominal obesity and glucose tolerance in obese adults: A randomized trial. *Annals of Internal Medicine*, *162*(5), 325–334. https://doi.org/10.7326/M14-1189

Rothman, A. (2000). Toward a Theory-Based Analysis of Behavioural Maintenance. *Health Psychology*, *19*(Sup1), 64–69. https://doi.org/10.1037/0278-6133.19.Suppl1.64

Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-09075-3

Rouault, M., & Fleming, S. M. (2020). Formation of global self-beliefs in the human brain. *Proceedings of the National Academy of Sciences*, 202003094. https://doi.org/10.1073/pnas.2003094117

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human Metacognition Across Domains: Insights from Individual Differences and Neuroimaging. *Personality Neuroscience*, *1*, E17. https://doi.org/10.1017/pen.2018.16

Rovio, S., Kåreholt, I., Helkala, E.-L., Viitanen, M., Winblad, B., Tuomilehto, J., … Kivipelto, M. (2005). Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease. *The Lancet Neurology*, *4*(11), 705–711. https://doi.org/10.1016/S1474-4422(05)70198-8

Ruscheweyh, R., Willemer, C., Krüger, K., Duning, T., Warnecke, T., Sommer, J., … Flöel, A. (2011). Physical activity and memory functions: An interventional study. *Neurobiology of Aging*, *32*(7), 1304–1319. https://doi.org/10.1016/j.neurobiolaging.2009.08.001

Ryan, R. M., & Deci, E. L. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist*, *55*(1), 68–78.

Ryan, R. M., Patrick, H., Deci, E. L., & Williams, G. C. (2008). Facilitating health behaviour change and its maintenance: Interventions based on Self-Determination Theory. *The European Health Psychologist*, *10*(1), 2–5.

Sackett, A. M., Wu, G., White, R. J., & Markle, A. B. (2015). *Harnessing optimism: How eliciting goals improves performance*. St. Paul, MN. https://doi.org/http://dx.doi.org/10.2139/ssrn.2544020

Salam, H., Marcora, S. M., & Hopker, J. G. (2017). The effect of mental fatigue on critical power during cycling exercise. *European Journal of Applied Physiology*, *118*(1), 85–92. https://doi.org/10.1007/s00421-017-3747-1

Saraç, S., & Karakelle, S. (2012). On-line and off-line assessment of metacognition. *International Electronic Journal of Elementary Education*, *4*(2), 301–315.

Sarmento, H., Clemente, F. M., Marques, A., Milanovic, Z., Harper, L. D., & Figueiredo, A. (2019). Recreational football is medicine against non-communicable diseases: A systematic review. *Scandinavian Journal of Medicine & Science in Sports*, *30*(4), 618–637. https://doi.org/https://doi.org/10.1111/sms.13611

Sbrocco, T., Nedegaard, R. C., Stone, J. M., & Lewis, E. L. (1999). Behavioral choice treatment promotes continuing weight loss: Preliminary results of a cognitive-behavioral decision-based treatment for obesity. *Journal of Consulting and Clinical Psychology*, *67*(2), 260–266. https://doi.org/10.1037/0022-006X.67.2.260

Schellings, G. (2011). Applying learning strategy questionnaires: problems and possibilities. *Metacognition and Learning*, *6*(2), 91–109. https://doi.org/10.1007/s11409-011-9069-5

Schellings, G., & Hout-Wolters, B. Van. (2011). Measuring strategy use with self-report instruments: theoretical and empirical considerations. *Metacognition and Learning*, *6*(2), 83–90. https://doi.org/10.1007/s11409-011-9081-9

Schellings, G. L. M., Hout-Wolters, B. H. A. M. van, Veenman, M. V. J., & Meijer, J. (2013). Assessing metacognitive activities : the in-depth comparison of a task-specific questionnaire with think-aloud protocols. *European Journal of Psychology of Education*, *28*, 963–990. https://doi.org/10.1007/s10212-012-0149-y

Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the “signal extraction” counterexplanation for the Dunning-Kruger effect in self-evaluation of performance. *Journal of Economic Psychology*, *39*, 85–100. https://doi.org/10.1016/j.joep.2013.07.004

Schmid, T. L., Jeffery, R. W., & Hellerstedt, W. L. (1989). Direct mail recruitment to home-based smoking and weight control programs: A comparison of strategies. *Preventive Medicine*, *18*(4), 503–517. https://doi.org/10.1016/0091-7435(89)90009-1

Schraw, G. (1997). The Effect of Generalized Metacognitive Performance Knowledge and Confidence. *The Journal of Experimental Education*, *65*(2), 135–146. https://doi.org/10.1080/00220973.1997.9943788

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Schraw, G., Kuch, F., & Gutierrez, A. P. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction*, *24*(1), 48–57. https://doi.org/10.1016/j.learninstruc.2012.08.007

Schraw, G., & Moshman, D. (1995). Metacognitive Theories. *Educational Psychology Review*, *7*(4), 351–371. https://doi.org/I040-726X/95/1200-0351507.50/0

Schumacher, N., Schmidt, M., Wellmann, K., & Braumann, K. M. (2018). General perceptual-cognitive abilities: Age and position in soccer. *PLoS ONE*, *13*(8), e0202627. https://doi.org/10.1371/journal.pone.0202627

Schunk, D. H., & DiBenedetto, M. K. (2016). Self-Efficacy Theory in Education. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of Motivation at School* (2nd ed, pp. 34–54). New York, NY: Routledge. https://doi.org/10.4324/9781315773384.ch3

Shepperd, J. A., & Mcnulty, J. K. (2002). The affective consequences of expected and unexpected outcomes. *Psychological Science*, *13*(1), 85–88. https://doi.org/https://doi.org/10.1111/1467-9280.00416

Sibley, B. A., & Bergman, S. M. (2018). What keeps athletes in the gym? Goals, psychological needs, and motivation of CrossFit™ participants. *International Journal of Sport and Exercise Psychology*, *16*(5), 555–574. https://doi.org/10.1080/1612197X.2017.1280835

Skorski, S., & Abbiss, C. R. (2017). The manipulation of pace within endurance sport. *Frontiers in Physiology*, *8*, 102. https://doi.org/10.3389/fphys.2017.00102

Smyth, B. (2018). Fast starters and slow finishers: A large-scale data analysis of pacing at the beginning and end of the marathon for recreational runners. *Journal of Sports Analytics*, *4*(3), 229–242. https://doi.org/10.3233/jsa-170205

Sobal, J., & Milgrim, M. (2019). Gendertyping sports: social representations of masculine, feminine, and neither-gendered sports among US university students. *Journal of Gender Studies*, *28*(1), 29–44. https://doi.org/10.1080/09589236.2017.1386094

Soderstrom, N. C., Mccabe, D. P., & Rhodes, M. G. (2012). Older Adults Predict More Recollective Experiences Than Younger Adults. *Psychology and Aging*, *27*(4), 1082–1088. https://doi.org/10.1037/a0029048

Sperling, R. A., Howard, B. C., Staley, R., & DuBois, N. (2004). Metacognition and Self-Regulated Learning Constructs. *Educational Research and Evaluation*, *10*(2), 117–139. https://doi.org/10.1076/edre.10.2.117.27905

Stone, N. J. (2000). Exploring the Relationship between Calibration and Self-Regulated Learning. *Educational Psychology Review*, *12*(4), 437–475. https://doi.org/10.1023/A:1009084430926

Suominen, T. H., Korhonen, M. T., Alén, M., Heinonen, A., Mero, A., Törmäkangas, T., & Suominen, H. (2017). Effects of a 20-week high-intensity strength and sprint training program on tibial bone structure and strength in middle-aged and older male sprint athletes: a randomized controlled trial. *Osteoporosis International*, *28*(9), 2663–2673. https://doi.org/10.1007/s00198-017-4107-z

Swain, P., Biggins, J., & Gordon, D. (2019). Marathon Pacing Ability: Training Characteristics and Previous Experience. *European Journal of Sport Science*. https://doi.org/10.1080/17461391.2019.1688396

Tarnopolsky, M. A. (2008). Sex Differences in Exercise Metabolism and the Role of 17-Beta Estradiol. *Medicine & Science in Sports & Exercise*, *40*(4), 648–654. https://doi.org/10.1249/MSS.0b013e31816212ff

Teixeira, P. J., Carraça, E. V, Markland, D., Silva, M. N., & Ryan, R. M. (2012). Exercise , physical activity, and self-determination theory: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *9*, 78. https://doi.org/https://doi.org/10.1186/1479-5868-9-78

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of Metacognitive Monitoring Affects Learning of Texts. *Journal of Educational Psychology*, *95*(1), 66–73. https://doi.org/10.1037/0022-0663.95.1.66

Thompson, P. D., Buchner, D., Piña, I. L., Balady, G. J., Williams, M. A., Marcus, B. H., … Wenger, N. K. (2003). Exercise and physical activity in the prevention and treatment of atherosclerotic cardiovascular disease: A statement from the council on clinical cardiology (subcommittee on exercise, rehabilitation, and prevention) and the council on nutrition, physical. *Circulation*, *107*(24), 3109–3116. https://doi.org/10.1161/01.CIR.0000075572.40158.77

Tobias, S., Everson, H. T., & Laitusis, V. (1999). Towards a Performance Based Measure of Metacognitive Knowledge Monitoring: Relationships with Self-Reports and Behavior Ratings. In *Research on Metacognitive Monitoring* (pp. 1–23). Montreal, Canada.

Toering, T. T., Elferink-Gemser, M. T., Jordet, G., & Visscher, C. (2009). Self- regulation and performance level of elite and non-elite youth soccer players. *Journal of Sport Sciences*, *27*(14), 1509–1517. https://doi.org/http://dx.doi.org/10.1080/02640410903369919

Trubee, N. W., Vanderburgh, P. M., Diestelkamp, W. S., & Jackson, K. J. (2014). Effects of heat stress and sex on pacing in marathon runners. *Journal of Strength and Conditioning Research*, *28*(6), 1673–1678. https://doi.org/10.1519/JSC.0000000000000295

Urban, K., & Urban, M. (2018). Influence of Fluid Intelligence on Accuracy of Metacognitive Monitoring in Preschool Children Fades with the Calibration Feedback. *Studia Psychologica*, *60*(2), 123–136. https://doi.org/10.21909/sp.2018.02.757

Van Cutsem, J., Marcora, S., De Pauw, K., Bailey, S., Meeusen, R., & Roelands, B. (2017). The Effects of Mental Fatigue on Physical Performance: A Systematic Review. *Sports Medicine*, *47*(8), 1569–1588. https://doi.org/10.1007/s40279-016-0672-0

Veenman, M. V. J. (2011). Alternative assessment of strategy use with self-report instruments: a discussion. *Metacognition and Learning*, *6*, 205–211. https://doi.org/10.1007/s11409-011-9080-x

Voss, M. W., Kramer, A. F., Basak, C., Prakash, R. S., & Roberts, B. (2010). Are expert athletes “Expert” in the cognitive laboratory? A meta-analytic review of cognition and sport expertise. *Applied Cognitive Psychology*, *24*(6), 812–826. https://doi.org/10.1002/acp.1588

Watson, S. L., Weeks, B. K., Weis, L. J., Harding, A. T., Horan, S. A., & Beck, B. R. (2018). High-Intensity Resistance and Impact Training Improves Bone Mineral Density and Physical Function in Postmenopausal Women With Osteopenia and Osteoporosis: The LIFTMOR Randomized Controlled Trial. *Journal of Bone and Mineral Research*, *33*(2), 211–220. https://doi.org/10.1002/jbmr.3284

Weinberg, B. A. (2009). A model of overconfidence. *Pacific Economic Review*, *14*(4), 502–515. https://doi.org/10.1111/j.1468-0106.2009.00466.x

Weisenthal, B. M., Beck, C. A., Maloney, M. D., DeHaven, K. E., & Giordano, B. D. (2014). Injury Rate and Patterns Among CrossFit Athletes. *Orthopaedic Journal of Sports Medicine*, *2*(4), 232596711453117. https://doi.org/10.1177/2325967114531177

Westcott, W. L. (2012). Resistance training is medicine: Effects of strength training on health. *Current Sports Medicine Reports*, *11*(4), 209–216. https://doi.org/10.1249/JSR.0b013e31825dabb8

Wiersma, R., Stoter, I. K., Visscher, C., Hettinga, F. J., & Elferink-Gemser, M. T. (2017). Development of 1500-m pacing behavior in junior speed skaters: A longitudinal study. *International Journal of Sports Physiology and Performance*, *12*(9), 1224–1231. https://doi.org/10.1123/ijspp.2016-0517

Windt, J., & Gabbett, T. J. (2016). How do training and competition workloads relate to injury? The workload — injury aetiology model. *British Journal of Sports Medicine*, *51*(5), 428–435. https://doi.org/10.1136/bjsports-2016-096040

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In *Handbook of Self-Regulation* (pp. 531–566). https://doi.org/https://doi.org/10.1016/B978-012109890-2/50045-7

World Health Organization. (n.d.). What is Moderate-intensity and Vigorous-intensity Physical Activity? Retrieved May 26, 2020, from https://www.who.int/dietphysicalactivity/physical_activity_intensity/en/

Wylie, S. A., Ally, B. A., van Wouwe, N. C., Neimat, J. S., van den Wildenberg, W. P. M., & Bashore, T. R. (2019). Exposing an "Intangible" Cognitive Skill Among Collegiate Football Players: III. Enhanced Reaction Control to Motion. *Frontiers in Sports and Active Living*, *1*, 51. https://doi.org/10.3389/fspor.2019.00051

Wylie, S. A., Bashore, T. R., Van Wouwe, N. C., Mason, E. J., John, K. D., Neimat, J. S., & Ally, B. A. (2018). Exposing an "Intangible" cognitive skill among collegiate football players: Enhanced interference control. *Frontiers in Psychology*, *9*, 49. https://doi.org/10.3389/fpsyg.2018.00049

Xiang, P., McBride, R. E., Lin, S., Gao, Z., & Francis, X. (2018). Students' Gender Stereotypes about Running in Schools. *Journal of Experimental Education*, *86*(2), 233–246. https://doi.org/10.1080/00220973.2016.1277335

Zepeda, C. D., Richey, J. E., Ronevich, P., & Nokes-Malach, T. J. (2015). Direct Instruction of Metacognition Benefits Adolescent Science Learning, Transfer, and Motivation: An In Vivo Study. *Journal of Educational Psychology*, *107*(4), 954–970. https://doi.org/http://dx.doi.org/10.1037/edu0000022 Direct

Zering, J. C., Brown, D. M. Y., Graham, J. D., & Bray, S. R. (2017). Cognitive control exertion leads to reductions in peak power output and as well as increased perceived exertion on a graded exercise test to exhaustion. *Journal of Sports Sciences*, *35*(18), 1799–1807. https://doi.org/10.1080/02640414.2016.1237777

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of Self-Regulation* (pp. 13–39). San Diego, CA. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183. https://doi.org/10.3102/000283120

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modeling*, *53*(1), 141–160.

CHAPTER 8: APPENDICES

8.1 ETHICAL APPROVAL LETTERS

8.1.1 Study 1

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

09 November 2018

Dear Kostas

Thank you for submitting your ethical application which was considered at the School of Psychology & Neuroscience Ethics Committee meeting on 1st November 2018; the following documents have been reviewed:

1. Ethical Application Form
2. Study Protocol and Study Flowchart
3. Participant Information and Questionnaire
4. Participant Debriefing Form
5. Data Management Plan

The School of Psychology & Neuroscience Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

Approval Code:	PS13950	**Approved on:**	01/11/2018	**Approval Expiry:**	01/11/2023
Project Title:	How accurate are runners in predicting their performance in 5km and 10km runs?				
Researcher:	Konstantinos Liverakos				
Supervisor:	Dr Akira O'Connor				

Approval is awarded for five years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the five year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (http://www.st-andrews.ac.uk/utrec/guidelinespolicies/). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Convener of the School Ethics Committee

cc Dr Akira O'Connor (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

8.1.2 Study 2

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

06 August 2018

Dear Konstantinos

Thank you for submitting your ethical application which was considered by the School of Psychology & Neuroscience Ethics Committee on 19th July 2018; the following documents have been reviewed:

1. Ethical Application Form
2. Study Protocol
3. Study Flowchart
4. Participant Debriefing Form
5. Questionnaire: Race Predictions
6. Data Management Plan

The School of Psychology & Neuroscience Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

Approval Code:	PS13876	**Approved on:**	02/08/2018	**Approval Expiry:**	02/08/2023
Project Title:	How does time of prediction influence running calibration in the Alloa Half Marathon?				
Researcher:	Konstantinos Liverakos				
Supervisor:	Dr Akira O'Connor				

Approval is awarded for five years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the five year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (http://www.st-andrews.ac.uk/utrec/guidelinespolicies/). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Convener of the School Ethics Committee

cc Dr Akira O'Connor (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

8.1.3 Studies 3 & 6

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

16 February 2018

Dear Konstantinos

Thank you for submitting your ethical application which was considered at the School of Psychology & Neuroscience Ethics Committee meeting on 8th February 2018; the following documents have been reviewed:

1. Ethical Application Form
2. Study Protocol and Flowchart
3. Advertisements (Poster, SONA, Facebook)
4. Participant Information Sheet
5. Participant Consent Form: Anonymous Data
6. Participant Debriefing Forms (Participating and Non-Participating)
7. External permission documentation (Functional Fitness St Andrews and University of St Andrews Sports Centre)
8. PAR-Q Health Eligibility Questionnaire and Exclusion Criteria
9. Experience Questionnaire
10. Risk Assessment
11. Data Management Plan

The School of Psychology & Neuroscience Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

Approval Code:	PS13328	**Approved on:**	15/02/2018	**Approval Expiry:**	15/02/2023
Project Title:	Investigating the effects of experience and gender on high-intensity functional movement exercise calibration				
Researcher:	Konstantinos Liverakos				
Supervisor:	Dr Akira O'Connor				

Approval is awarded for five years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the five year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Cont.

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (http://www.st-andrews.ac.uk/utrec/guidelinespolicies/). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Convener of the School Ethics Committee

cc Dr Akira O'Connor (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

8.1.4 Study 4

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

11 September 2018

Dear Konstantinos

Thank you for submitting your ethical application which was considered by the School of Psychology & Neuroscience Ethics Committee on 31st August; the following documents have been reviewed:

1. Ethical Application Form
2. Study Protocol
3. Study Flowchart
4. Advertisements: Poster, SONA, Facebook
5. Participant Information Sheet
6. Participant Consent Form: Anonymous Data
7. Participant Debriefing Form and No Participation Debriefing Form
8. Questionnaires: PAR-Q Health Eligibility Form (and PAR-Q Health Eligibility Form Appropriate Responses information), Metacognitive Awareness Inventory, Metacognitive Awareness Inventory for Exercise, Behavioural Regulation in Exercise Questionnaire - 3
9. Data Management Plan
10. Risk Assessment

The School of Psychology & Neuroscience Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

Approval Code:	PS13905	**Approved on:**	06/09/2018	**Approval Expiry:**	06/09/2023
Project Title:	To what extent do academic and exercise self-regulation influence exercise calibration and motivation?				
Researcher:	Konstantinos Liverakos				
Supervisor:	Dr Akira O'Connor				

Approval is awarded for five years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the five year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Cont.

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (http://www.st-andrews.ac.uk/utrec/guidelinespolicies/). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Convener of the School Ethics Committee

cc Dr Akira O'Connor (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

8.1.5 Studies 5 & 8

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

27 June 2019

Dear Kostas

Thank you for submitting your ethical application which was considered at the School of Psychology & Neuroscience Ethics Committee meeting on 20 June 2019; the following documents have been reviewed:

1. Ethical Application Form
2. Study Protocol and Study Flowchart
3. Advertisement: Poster, Facebook and SONA versions
4. Participant Information Sheet
5. Participant Consent Form
6. Participant Debrief: Participation and No Participation versions
7. Questionnaires: PAR-Q Health Eligibility Form with PAR-Q Eligibility Appropriate Responses Form, Exercise experience questionnaire, Metacognitive Awareness Inventory for Running
8. External Permission: University of St Andrews Sports Centre

The School of Psychology & Neuroscience Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

Approval Code:	PS14429	**Approved on:**	26/06/2019	**Approval Expiry:**	26/06/2024
Project Title:	Examining the effects of prediction guidance priming and self-regulation on running calibration accuracy				
Researcher:	Konstantinos Liverakos				
Supervisor:	Dr Akira O'Connor				

Approval is awarded for five years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the five-year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Cont.

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (http://www.st-andrews.ac.uk/utrec/guidelinespolicies/). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Convener of the School Ethics Committee

Cc Dr Akira O'Connor (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

8.1.6 Study 7

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

14 February 2019

Dear Kostas

Thank you for submitting your ethical application which was considered at the School of Psychology & Neuroscience Ethics Committee meeting on 7th February 2019; the following documents have been reviewed:

1. Ethical Application Form
2. Protocol and Study Flowchart
3. Advertisements: Poster, Facebook, SONA and Student Memos
4. Participant Information Sheet
5. Participant Consent Form
6. Debrief: Participant and No-Participation versions
7. Questionnaires

The School of Psychology & Neuroscience Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

Approval Code:	PS14081	**Approved on:**	07/02/2019	**Approval Expiry:**	07/02/2024
Project Title:	How does prediction guidance influence calibration accuracy in high-intensity functional movement exercise?				
Researcher:	Konstantinos Liverakos				
Supervisor:	Dr Akira O'Connor				

Approval is awarded for five years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the five year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (http://www.st-andrews.ac.uk/utrec/guidelinespolicies/). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Convener of the School Ethics Committee cc Dr Akira O'Connor (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

8.2 STUDY 1 QUESTIONNAIRE

Information

Saturday / Sunday

The questions below are part of a research study being conducted by researchers in the School of Psychology & Neuroscience at the University of St Andrews. Your participation in the study will in no way affect your registration and participation in the Edinburgh Christmas events. If you do not wish to participate in the study, please just return the form to the experimenter. If you do not wish to answer a question, skip that question and proceed to the next. The data you provide will be attributable to you until it is matched to your event result. At this point, we will anonymise your data and store it in a non-identifiable format.

Demographic Data

The questions below examine factors that influence the accuracy of predictions you make about your running.

1. I am participating in the (Please circle as appropriate) **5km / 10km** run.

2. **Name** (as used at Edinburgh Christmas Run registration; BLOCK CAPITALS):

____

3. Date of Birth: DD / MM / YY

4. I am a member of a running club (Please circle as appropriate): **YES** / **NO**

5. I have run or jogged for exercise for _______ **years**, and _______ **months**.

6. In an average week, I run or jog approximately _______ **miles OR** _______ **km**.

7. Over the past two months, I have consistently engaged in the following types of training (tick all that apply):

Interval training	☐	Hill training	☐
Tempo runs	☐	Long runs	☐

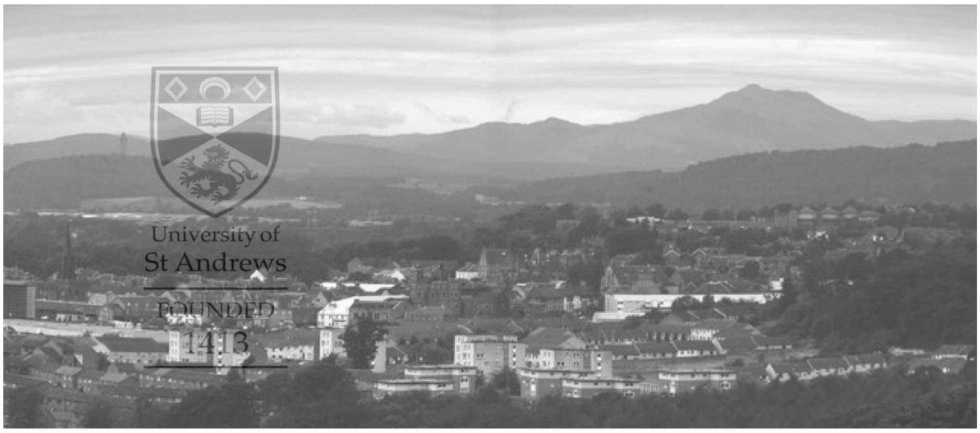
Predictions

Please read all of the questions below before completing this section. For each question, give your best guess of an exact time in minutes and seconds (and not a time window or range) for your Christmas run.

1. The finish time I **hope to achieve** (my goal time) is MM : SS

2. The finish time I **think is most likely for me to achieve** is MM : SS

8.3 STUDY 2 QUESTIONNAIRE



ALLOA HALF MARATHON STUDY QUESTIONNAIRE

The questions below examine the factors that influence your running predictions.

*: Required

1. First name (as used at Alloa Half Marathon registration):* [FIRSTNAME]

2. Middle name (Optional): [MIDDLENAME]

3. Last name (as used at Alloa Half Marathon registration):* [LASTNAME]

4. Date of Birth:* [DD] / [MM] / [YYYY]

5. I have run or jogged for [ ] years*, and [ ] months*.

6. Have you participated in a Half Marathon before?*
   Select...

7. If you answered NO to question 6, please ignore this question.
   If you answered YES, please indicate your last half marathon finish time:
   Select... hours and: Select... minutes.

8. In an average week, I run or jog approximately ____ miles, **OR** ____ km.

9. Over the past two months, I have consistently engaged in the following types of training (tick all that apply):
   - Interval training ☐
   - Hill training ☐
   - Tempo runs ☐
   - Long runs ☐

10. Date when questions were answered: 03/08/2020

Predictions

Please read both of the questions below before completing this section. For each question, give your best guess of an exact time in hours and minutes.

1. The finish time I hope to achieve (my goal time) is:*
   Select... hours and:* Select... minutes.

2. The finish time I think is most likely for me to achieve is:* Select... hours and:* Select... minutes.

Submit

If you have any problems or questions, please email Konstantinos Liverakos: **kl65@st-andrews.ac.uk**

8.4 BORG EXERTION SCALE

20-Grade Scale	
6	
7	Very, very light
8	
9	Very light
10	
11	Fairly light
12	
13	Somewhat hard
14	
15	Hard
16	
17	Very hard
18	
19	Very, very hard
20	

Scale for perceived exertion rating. Figure based on Borg (1982).

8.5 PAR-Q HEALTH ELIGIBILITY FORMS

8.5.1 PAR-Q Forms for Studies 3 & 6

Please provide an answer to each of the following questions assessing your health eligibility to participate in the study:

1. Has your doctor ever said that you have a bone or joint problems, such as arthritis that has been aggravated by exercise or might be made worse with exercise? **YES / NO**

2. Do you have high blood pressure? **YES / NO**

3. Do you have low blood pressure? **YES / NO**

4. Do you have Diabetes Mellitus or any other metabolic disease? **YES / NO**

5. Has your doctor ever said you have raised cholesterol (serum level above 6.2mmol/L)? **YES / NO**

6. Has your doctor ever said that you have a heart condition arid that you should only do physical activity recommended by a doctor? **YES / NO**

7. Have you ever felt pain in your chest when you do physical exercise? **YES / NO**

8. Is your doctor currently prescribing you drugs or medication? **YES / NO**

9. Have you ever suffered from unusual shortness of breath at rest or with mild exertion? **YES / NO**

10. Is there any history of Coronary Heart Disease in your family? **YES / NO**

11. Do you often feel faint, have spells of severe dizziness or have lost consciousness? **YES / NO**

12. Do you currently drink more than the average amount of alcohol per week (21 units for men and 14 units for women)? **YES / NO**

13. Do you currently smoke? **YES / NO**

14. Are you, or is there any possibility that you might be pregnant? **YES / NO**

15. Do you know of any other reason why you should not participate in a physical activity programme? **YES / NO**

If you answered YES to any of the questions above please give details:

If you answered YES to one or more questions: For safety reasons you will not be eligible to participate in the study, unless the details you have provided indicate a lack of potential danger.

If you answered NO to all questions: If you answered PAR-Q accurately, you have reasonable assurance of your present suitability for participating in a maximal exertion task.

Assumption of Risk

I hereby state that I have read, understood and answered honestly the questions above. I also state that I wish to participate in activities, which will include strenuous aerobic exercise. I realise that my participation in these activities involves the risk of injury and exhaustion. Furthermore, I hereby confirm that I am voluntarily engaging in a maximal level of exercise, which has been set to me by the researcher.

Participant 's Signature : **Date :**

8.5.2 PAR-Q Forms for Studies 4, 5, 7 & 8

Please provide an answer to each of the following questions assessing your health eligibility to participate in the study:

1. Has your doctor ever said that you have a bone or joint problems, such as arthritis that has been aggravated by exercise or might be made worse with exercise? **YES / NO**

2. Do you have high blood pressure? **YES / NO**

3. Do you have low blood pressure? **YES / NO**

4. Do you have Diabetes Mellitus or any other metabolic disease? **YES / NO**

5. Has your doctor ever said you have raised cholesterol (serum level above 6.2mmol/L)? **YES / NO**

6. Has your doctor ever said that you have a heart condition arid that you should only do physical activity recommended by a doctor? **YES / NO**

7. Have you ever felt pain in your chest when you do physical exercise? **YES / NO**

8. In the past month, have you felt pain in your chest when you were not doing physical activity? **YES / NO**

9. Is your doctor currently prescribing you drugs or medication? **YES / NO**

10. Have you ever suffered from unusual shortness of breath at rest or with mild exertion? **YES / NO**

11. Is there any history of Coronary Heart Disease in your family? **YES / NO**

12. Do you often feel faint, have spells of severe dizziness or lose consciousness? **YES / NO**

13. Do you currently drink more than the average amount of alcohol per week (21 units for men and 14 units for women)? **YES / NO**

14. Do you currently smoke? **YES / NO**

15. Are you, or is there any possibility that you might be pregnant? **YES / NO**

16. Do you know of any other reason why you should not participate in physical activity? **YES / NO**

If you answered YES to any of the questions above please give details:

If you answered YES to one or more questions: For safety reasons you will not be eligible to participate in the study, unless the details you have provided indicate that you are not at risk of experiencing adverse effects (e.g. faint or experience discomfort) during or following exercise.

If you answered NO to all questions: If you answered PAR-Q accurately, you have reasonable assurance of your present suitability for participating in a maximal exertion task.

Assumption of Risk
I hereby state that I have read, understood and answered honestly the questions above. I also state that I wish to participate in activities, which will include strenuous aerobic exercise. I realise that my participation in these activities involves the risk of injury and exhaustion. Furthermore, I hereby confirm that I am voluntarily engaging in a maximal level of exercise, which has been set to me by the researcher.

Participant's Signature: **Date:**

8.6 HIFME EXPERIENCE QUESTIONNAIRE

Sex assigned at birth: **Male** **Female** **Neither** **Prefer Not to Say**

Age: ____________

For how long have you been participating in high-intensity functional movement exercise: _________ years ____________ months

What form of high-intensity functional movement exercise do you take part in? (e.g. CrossFit, body sculpting, circuit training, etc.)

____
____
____
____

How frequently do you participate in high-intensity functional movement exercise: _________times per week

When was your last workout?

____

Do you take part in other physical activities?

Activity 1: ____________________________

For how long: ______ years ______ months **How frequently: ______ times per week**

Activity 2: ____________________________

For how long: ______ years ______ months **How frequently: ______ times per week**

Activity 3: ____________________________

For how long: ______ years ______ months **How frequently: ______ times per week**

Have you participated in high-intensity functional movement exercise competitions? YES/NO

If yes, provide details (competition name, date):

8.7 METACOGNITIVE AWARENESS INVENTORY (MAI)

Think of yourself as a learner. Read each statement carefully. Consider if the statement applies to you when you are in the role of a learner (student, attending classes, university, etc.). To indicate whether you agree or disagree with the statement, circle the appropriate value from 1 to 5 (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree).

1. I ask myself periodically if I am meeting my goals.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

2. I consider several alternatives to a problem before I answer.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

3. I try to use strategies that have worked in the past.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

4. I pace myself while learning in order to have enough time.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

5. I understand my intellectual strengths and weaknesses.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

6. I think about what I really need to learn before I begin a task

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

7. I know how well I did once I finish a test.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

8. I set specific goals before I begin a task.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

9. I slow down when I encounter important information.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

10. I know what kind of information is most important to learn.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

11. I ask myself if I have considered all options when solving a problem.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

12. I am good at organizing information.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

13. I consciously focus my attention on important information.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

14. I have a specific purpose for each strategy I use.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

15. I learn best when I know something about the topic.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

16. I know what the teacher expects me to learn.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

17. I am good at remembering information.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

18. I use different learning strategies depending on the situation.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

19. I ask myself if there was an easier way to do things after I finish a task.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

20. I have control over how well I learn.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

21. I periodically review to help me understand important relationships.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

22. I ask myself questions about the material before I begin.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

23. I think of several ways to solve a problem and choose the best one.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

24. I summarize what I’ve learned after I finish.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

25. I ask others for help when I don’t understand something.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

26. I can motivate myself to learn when I need to

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

27. I am aware of what strategies I use when I study.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

28. I find myself analysing the usefulness of strategies while I study.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

29. I use my intellectual strengths to compensate for my weaknesses.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

30. I focus on the meaning and significance of new information.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

31. I create my own examples to make information more meaningful.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

32. I am a good judge of how well I understand something.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

33. I find myself using helpful learning strategies automatically.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

34. I find myself pausing regularly to check my comprehension.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

35. I know when each strategy I use will be most effective.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

36. I ask myself how well I accomplish my goals once I'm finished.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

37. I draw pictures or diagrams to help me understand while learning.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

38. I ask myself if I have considered all options after I solve a problem.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

39. I try to translate new information into my own words.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

40. I change strategies when I fail to understand.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

41. I use the organizational structure of the text to help me learn.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

42. I read instructions carefully before I begin a task.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

43. I ask myself if what I'm reading is related to what I already know.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

44. I re-evaluate my assumptions when I get confused.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

45. I organize my time to best accomplish my goals.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

46. I learn more when I am interested in the topic.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

47. I try to break studying down into smaller steps.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

48. I focus on overall meaning rather than specifics.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

49. I ask myself questions about how well I am doing while I am learning something new.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

50. I ask myself if I learned as much as I could have once I finish a task.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

51. I stop and go back over new information that is not clear.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

52. I stop and reread when I get confused.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

8.8 METACOGNITIVE AWARENESS INVENTORY FOR EXERCISE (MAIE)

Think of yourself as an athlete (recreational or high level). Read each statement carefully. Consider if the statement applies to you when you are in the role of an athlete (exercising, training, playing a sport, competing, being part of a sports team, etc.). To indicate whether you agree or disagree with the statement, circle the appropriate value from 1 to 5 (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree).

1. I focus more on my technique and/or performance when I perform important and challenging exercises.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

2. I plan on a long-term exercise programme in order to have enough time to achieve my performance goals.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

3. I perform best when I have experience with the sport/exercise.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

4. I try to use strategies that have worked for workouts/competitions/matches in the past.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

5. I understand my strengths and weaknesses in sports/exercise.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

6. I consciously focus my attention on important exercise technique/information.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

7. I know how well I did once I finish a competition/training session.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

8. I think about how I need to perform before I begin a workout.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

9. I ask myself questions about how well I am doing while exercising.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

10. I know what kind of exercise is most important to practice.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

11. I use different training strategies depending on the situation.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

12. I ask myself if there was a better way to do things after I finish a task.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

13. I ask others for help/advice when I don't perform well.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

14. I am aware of the types of exercise I engage in when I train.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

15. I am good at organising my exercise programme.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

16. I set specific goals before I begin a workout.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

17. I focus on the execution and significance of new exercises.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

18. I consider several alternatives on how to tackle a workout before I begin.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

19. I change strategies when I fail to perform in the expected way.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

20. I summarise my improvements after I finish a training programme.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

21. I ask myself periodically if I am meeting my goals.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

22. I take videos whilst I'm exercising to help me improve my form.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

23. I ask myself how well I have accomplished my goals once I'm finished.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

24. I re-evaluate my assumptions when I do not perform well.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

25. I ask myself questions about the workout/training session before I begin.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

26. I find myself using helpful exercise techniques automatically.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

27. I am good at performing in sports/exercise.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

28. I use my fitness/strength when I need to compensate for my weaknesses in technique.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

29. I try to adjust new exercises for my own body characteristics.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

30. I ask myself if I have considered all strategies (e.g. pacing in running, weight lifted and rest periods during weightlifting, tactics in sports such as football, etc.) before I attempt a workout.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

31. I stop and go back over technique steps when my form fails.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

32. I think of several ways to perform a workout and choose the best one.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

33. I have control over how well I perform in sport/exercise.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

34. I use the organizational structure of a workout to help me understand how to improve my performance/fitness.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

35. I know when each strategy I use will be most effective.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

36. I ask myself if I have considered all options after I plan my strategy for a workout.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

37. I periodically review strategies to help me understand how to best improve my performance.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

38. I am a good judge of how well I perform an exercise (e.g. technique).

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

39. I make sure I know what I'm going to do before I begin an exercise.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

40. I ask myself if the exercises I'm practicing are related to what I can already do.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

41. I find myself evaluating why I am doing a particular exercise as part of a training programme.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

42. I ask myself if I challenged myself as much I could have once I finish a workout.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

43. I stop and rethink about the exercise technique when I feel that my form is not good/satisfactory.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

44. I perform better when I am interested in the sport/exercise.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

45. I organise my time and effort to best accomplish my exercise goals.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

46. I try to break workouts down into smaller steps.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

47. I find myself analysing the usefulness/effectiveness of strategies I use while I exercise.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

48. I can motivate myself to train when I need to.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

49. I know how I am expected to perform in sport/exercise.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

50. I have a specific purpose for each type of exercise I use when I train.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

8.9 METACOGNITIVE AWARENESS INVENTORY FOR RUNNING (MAIR)

Think of yourself as a runner (recreational or high level). Read each statement carefully. Consider if the statement applies to you as a runner (exercising, training, competing, being part of a running team, etc.). If a statement does not specifically mention running, then answer in terms of general exercise. To indicate whether you agree or disagree with the statement, circle the appropriate value from 1 to 5 (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree).

1. I focus more on my technique and/or performance when I perform important and challenging exercises.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

2. I plan on a long-term running programme in order to have enough time to achieve my running performance goals.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

3. I perform best when I have experience with the sport/exercise.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

4. I try to use strategies that have worked for runs/competitions in the past.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

5. I understand my strengths and weaknesses in running.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

6. I consciously focus my attention on important running technique/information.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

7. I know how well I did once I finish a competition/training session.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

8. I think about how I need to perform before I begin a run.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

9. I ask myself questions about how well I am doing while running.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

10. I know what kind of running training is most important to practice.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

11. I use different running strategies depending on the situation/goals.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

12. I ask myself if there was a better way to do things after I finish a run/race.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

13. I ask others for running advice when I don't perform well.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

14. I am aware of the types of exercise I engage in when I train.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

15. I am good at organising my running programme.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

16. I set specific goals before I begin a run/race.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

17. I focus on the execution and significance of new exercises.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

18. I consider several alternatives on how to tackle a run/race before I begin.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

19. I change strategies when I fail to perform in the expected way in running.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

20. I summarise my improvements after I finish a running training programme.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

21. I ask myself periodically if I am meeting my running goals.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

22. I take videos whilst I'm exercising to help me improve my form.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

23. I ask myself how well I have accomplished my goals once I'm finished running.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

24. I re-evaluate my assumptions when I do not run well.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

25. I ask myself questions about the run/race before I begin.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

26. I find myself using helpful running techniques automatically.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

27. I am good at running.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

28. I use my fitness/strength when I need to compensate for my weaknesses in technique.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

29. I try to adjust new exercises for my own body characteristics.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

30. I ask myself if I have considered all strategies (e.g. pacing) before I attempt a run/race.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

31. I stop and go back over technique steps when my form fails.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

32. I think of several ways to perform a workout and choose the best one.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

33. I have control over how well I run.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

34. I use the organizational structure of a workout to help me understand how to improve my performance/fitness.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

35. I know when each running strategy I use will be most effective.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

36. I ask myself if I have considered all options after I plan my strategy for a run/race.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

37. I periodically review strategies to help me understand how to best improve my running.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

38. I am a good judge of how well I perform an exercise (e.g. technique).

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

39. I make sure I know what I'm going to do before I begin a run/race.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

40. I ask myself if the exercises I'm practicing are related to what I can already do.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

41. I find myself evaluating why I am doing a particular exercise as part of a training programme.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

42. I ask myself if I challenged myself as much I could have once I finish a run/race.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

43. I stop and rethink about the exercise technique when I feel that my form is not good/satisfactory.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

44. I perform better when I am interested in the sport/exercise.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

45. I organise my time and effort to best accomplish my running goals.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

46. I try to break workouts down into smaller steps.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

47. I find myself analysing the usefulness/effectiveness of strategies I use while I run.

1	2	3	4	5
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

48. I can motivate myself to run when I need to.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

49. I know how I am expected to perform in running.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

50. I have a specific purpose for each type of exercise I use when I train.

1	**2**	**3**	**4**	**5**
Strongly disagree	Sometimes disagree	Neutral	Sometimes agree	Strongly Agree

8.10 RUNNING EXPERIENCE QUESTIONNAIRE

Age: _______ years

Gender: Male Female Non-binary/third gender Prefer not to say

Prefer to self-describe _____________

The following questions have been developed to examine your previous exercise background. Please read and answer each question by writing your answer in the designated space or by circling the option that applies to you. If you have any questions, please ask the researcher for clarifications.

1. Years of running experience **after turning 18** (as an individual sport and not just as part of another sport, e.g. football): __________________________
2. Years of competitive running experience: _____________
3. Do you have other exercise experience over the past 5 years? **YES / NO**
4. If yes, what is your other exercise experience? _____________
5. Are you a member of a running club? **YES / NO**
6. How many kilometres do you usually run per week? (over the past 2 months): _____________
7. Do you ever run on a treadmill? **YES / NO**
8. If yes, do you usually run on a treadmill or outdoors? _____________
9. How would you rate your running training intensity (over the past 2 months)? **LOW / MEDIUM / HIGH**
10. Do you normally train for short (<1km), medium (1km to 10km), or long (>10km) distances? **SHORT / MEDIUM / LONG**
11. What was the distance of the last race in which you competed? _____________
12. What was your finish time? _____________
13. Do you run independently or do you receive feedback/supervision by a coach? **INDEPENDENTLY / COACH SUPERVISION (FEEDBACK)**
14. Do you usually run by yourself or with others? **MYSELF / OTHERS**