

Running head: mental toughness and behavioural perseverance

**Mental Toughness and Behavioural Perseverance: A Conceptual Replication and
Extension**

Brandon Giles¹, Paul Goods², Dylan Warner¹, Dale Quain¹, Peter Peeling^{1,2}, Kagan Ducker³,
Brian Dawson¹, and *Daniel F. Gucciardi³

¹*School of Sport Science, Exercise and Health, The University of Western Australia*

²*Western Australian Institute of Sport*

³*School of Physiotherapy and Exercise Science, Curtin University*

Author Notes

*Address correspondence to Daniel Gucciardi, School of Physiotherapy and Exercise Science, Curtin University, GPO Box U1987, Western Australia, 6845. Phone: +61 8 9266 3653. Email: daniel.f.gucciardi@gmail.com. Daniel Gucciardi is supported by a Curtin Research Fellowship.

Accepted for publication in *Journal of Science and Medicine in Sport* on October 31st 2017.

Abstract

Objectives: The purpose of this study was to conduct a conceptual replication of the proposition that mental toughness is associated positively with behavioural perseverance.

Design: Repeated-measures design.

Methods: In total, 38 male Australian rules footballers took part in this study (age, 21 ± 3 y; mass, 82.7 ± 11.0 kg; height, $1.84 \pm .07$ m; football experience, 13 ± 4 y). Participants self-reported mental toughness approximately one week prior to their first testing session where we assessed their aerobic capacity via the measurement of peak oxygen consumption ($\dot{V}O_{2\text{peak}}$). Approximately one week later, participants completed a 20m shuttle run test (MST). The final testing session took place approximately one week later, where participants completed a simulated team game circuit (STGC; 60 min) to simulate game-relevant level of fatigue, which was followed immediately by a 20m MST.

Results: Mental toughness was a salient determinant of the variation in behavioural perseverance under typical circumstances, when prior knowledge from past research was incorporated directly into the estimation process. However, the positive association between mental toughness and behavioural perseverance did not generalise to a performance context in which participants were fatigued.

Conclusions: The results of the current study suggest that mental toughness represents a salient psychological correlate of behavioural perseverance in a discrete physical task that taxes the aerobic energy system in some but not all situations. When fatigued, the effect of mental toughness is outweighed by greater underlying fitness.

Keywords: aerobic capacity; athlete; Bayesian; multistage shuttle test; team sport

Introduction

The concept of mental toughness has garnered substantial interest from researchers, practitioners and the general public over the past two decades. Within the academic literature, researchers have proposed several unique definitions and conceptual models of mental toughness¹. Although these definitions and models differ in their breadth (e.g., unidimensional versus multidimensional structure) and content (e.g., resources considered characteristic of the concept), they share commonality in terms of an overarching conceptualisation of mental toughness as a psychological capacity of individuals that characterises their potential for action towards an objective or purpose. Informed by this conceptual commonality, mental toughness has been defined recently¹ as “a state-like psychological resource that is purposeful, flexible, and efficient in nature for the enactment and maintenance of goal-directed pursuits” (p. 18). Aligned with this definition, there is converging evidence to support the adaptive nature of mental toughness for performance and well-being across a range of performance settings, including sport, education and the workplace^{1,2}.

In testing empirically the conceptual features of mental toughness, researchers have honed in on behavioural perseverance as a key behaviour by which this psychological capacity translates potential for action into goal attainment, high performance and well-being. In a sample of adolescent elite cricketers, Bell and colleagues³ examined the effect of an intervention to enhance mental toughness via punishment-conditioned stimuli (i.e., consequences for behaviour), administered within a transformational leadership climate in which a higher-order vision and purpose transcends short-term goals. Those players who received the intervention (n = 21) improved their mental toughness over a 12-month period when compared with a control group (n = 20). In turn, these improvements in mental toughness were associated with differences in behavioural perseverance (as indexed by

enhanced performance in the 20-m multistage shuttle run test [MST]) between the intervention and control groups 12-months after receiving the intervention ($d = 1.20$). Gucciardi and colleagues⁴ extended this work by examining the link between mental toughness and behavioural perseverance in a field setting, using a sample of adolescent Australian footballers ($N=330$). Adopting a cross-sectional design, they found that self-reported mental toughness was a salient correlate of MST performance ($\beta = .24$). Collectively, these two studies provide preliminary evidence to support this conceptual proposition of mental toughness (also see Gucciardi et al⁵, for evidence in educational and workplace settings).

Against this empirical backdrop, this study was designed as a conceptual replication because we aimed to test the theoretical proposition that mental toughness is associated positively with behavioural perseverance using an experimental protocol that addressed methodological shortcomings of this past work^{3,4}. First, we incorporated an objective measure of participants' aerobic capacity to examine the salience of mental toughness for behavioural perseverance while accounting for their physical fitness. Second, we separated temporally the assessment of mental toughness and all physical performance tests to minimise concerns associated with a mere measurement effect (i.e., enhanced accessibility of mental toughness for subsequent behaviour⁶). Third, all testing sessions were held at the same time of day to avoid any influence of circadian variation on performance⁷. As such, the first research question in this study sought to examine the extent to which these past findings replicate with a protocol that alleviates methodological concerns of previous research (RQ1).

In addition to this conceptual replication, a second purpose of the current study was to extend past work^{3,4} by testing the robustness of the association between mental toughness and behavioural perseverance under circumstances when participants are physically fatigued. This methodological extension speaks to a core proposition reported by athletes and coaches in

past qualitative work, namely that athletes with high levels of mental toughness are able to ‘push through the pain barrier’ when their body is telling them to give up^{8,9,10}. In other words, mental toughness may be most important when the ‘going gets tough’. This conceptual proposition has not yet been tested, and therefore, it remains unknown whether or not this post-hoc explanation of mental toughness by athletes and coaches extends to work conducted in the field. As such, the second research question in this study (RQ2) sought to test the degree to which mental toughness explains the variation in behavioural perseverance under circumstances when participants are fatigued, while accounting for baseline aerobic capacity and typical MST performance.

Methods

Monte Carlo simulations¹¹ indicated that at least 30 participants would provide sufficient power (>80%) to detect the expected effects of aerobic capacity and mental toughness on behavioural perseverance when incorporating prior beliefs directly into the estimation process (see Table 1); full details of these analyses are provided in the supplementary material. In total, 38 male Australian rules footballers provided informed consent and took part in this study (age, 21.18 ± 2.82 y; mass, 82.67 ± 11.02 kg; height, $1.84 \pm .07$ m; football experience, 12.89 ± 4.42 y). Footballers were recruited from clubs within the West Australian Football League (WAFL) and Western Australian Amateur Football League (WAAFL). All players were injury free and engaged in regular training at the time of data collection. The study protocol received approval from the lead author’s university ethics committee before participant recruitment.

Using a repeated-measures design, participants completed three testing sessions over a two-week period; each testing session was separated by one week. Approximately one week prior to attending the lab in session one, players self-reported their mental toughness⁵ via an online questionnaire (Qualtrics LLC, Utah, USA); participants rated the extent to which eight

items were reflective of how they typically thought, felt and behaved as a footballer (e.g., “I strive for continued success” and “I am able to regulate my focus when performing tasks”). Responses were recorded using a 7-point scale (1 = false, 100% of the time to 7 = true, 100% of the time). A total score was created by averaging participants’ responses across these eight items ($\alpha = .73$). In session one, participants completed a laboratory-based graded exercise test (GXT) to assess aerobic capacity via the measurement of peak oxygen consumption ($\dot{V}O_{2\text{peak}}$). In session two, footballers completed a stand-alone 20 m MST in an indoor gymnasium on a sprung wooden floor. Finally, in session three, participants first completed a 60 min simulated team game circuit (STGC)¹² on an outdoor grass oval, followed immediately by a MST, again completed in an indoor gymnasium on a sprung wooden floor. The STGC rather than a real match scenario was employed for pragmatic and scientific reasons (i.e., experimental control over the fatiguing protocol). All testing sessions were held at the same time of day to avoid any influence of circadian variation on performance. MST scores (shuttle:level) were converted to a decimal score for ease of analysis and reporting. Fluid intake was restricted to 600 ml of water within each session, in order to control for the influence of consumption variations; for the STGC, participants received a total of 600 ml of Gatorade (carbohydrate: 36 g, energy intake: 618 kJ; Schweppes Australia) consumed during the quarter and half-time breaks interspersed throughout the session. Full details of the experimental protocol and individual tests are provided in the online supplementary material.

The primary questions were tested with Bayesian regression analysis in *Mplus* 8¹³ using Markov Chain Monte Carlo (MCMC) simulation procedures with a Gibbs sampler. Within the context of Bayesian estimation, MCMC algorithms “mix” prior beliefs with observed data to produce “an approximation of the joint distribution of all parameters” in a model (¹⁴, p. 334; for a technical discussion, see¹⁵). As such, the posterior distribution represents the combination of prior beliefs and the new data at hand. We specified four

Markov chains each with a fixed number of 100,000 iterations to describe the posterior distribution, where the initial 50,000 iterations represented the burn-in phase. Model convergence was assessed using statistical (i.e., potential scale reduction factor; $PSR < 1.05$) and visual criteria (i.e., inspection of trace plots for stability in mean and variance of each chain). Posterior predictive checking is used to assess model fit in Bayesian estimation, where the posterior distribution is compared with the observed data to examine the degree to which the replicated data matches the observed data¹⁴. The posterior predictive *p*-value (PPP) and associated 95% credibility interval (CI) is produced in *Mplus*; PPP values close to .50 and a symmetric 95% CI centring on zero reflect an excellent fitting model, though typically values greater than .05 are considered acceptable¹⁴. Parameter estimates were considered credible when the 95% CI excluded zero. An overview of the priors employed for each research question are detailed in Table 1. Readers are referred elsewhere for accessible overviews of Bayesian analysis^{16,17}. For RQ1, $\dot{V}O_{2peak}$ and mental toughness were included as predictors of MST performance (decimal). For RQ2, footballers' baseline MST score was added alongside $\dot{V}O_{2peak}$ and mental toughness as a predictor of MST performance when fatigued. All variables were standardised prior to analysis to provide a common metric¹⁴.

Results

Validity evidence for the utility of the STGC protocol is provided in the supplementary material. $\dot{V}O_{2peak}$ ranged between 49.0 and 67.9 $\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ (60.5 ± 4.6 $\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$); baseline MST scores ranged from 10.45 to 14.08 (12.48 ± 0.99); and MST scores when fatigued ranged from 8.73 to 13.62 (11.14 ± 1.36). With regard to RQ1, the probability of the hypothesised model, given the data, was acceptable (PPP = .515, $\Delta_{\text{observed and replicated}} \chi^2$ 95% CI [-12.87, 12.32]). Visual inspection of trace plots and an examination of the PSR development over iterations (i.e., started at 1.006, reduced to 1 after 1800 iterations and remained at 1 throughout the full sequence) verified support for model

convergence¹⁴. $\dot{V}O_{2\text{peak}}$ ($\beta = .57$; 95% CI = .41, .73) and mental toughness ($\beta = .22$; 95% CI = .10, .35) were credible predictors, accounting for 34% of the variance in baseline MST performance. The correlation between $\dot{V}O_{2\text{peak}}$ and mental toughness was inverse and small ($r = -.04$; 95% CI = -.51, .40).

With regard to RQ2, the probability of the hypothesised model, given the data, was acceptable (PPP = .111, Δ observed and replicated χ^2 95% CI [-6.94, 28.77]). Visual inspection of trace plots and an examination of the PSR development over iterations (i.e., started at 1.007, reduced to 1 after 2100 iterations and remained at 1 throughout the full sequence) verified support for model convergence¹⁴. $\dot{V}O_{2\text{peak}}$ ($\beta = .40$; 95% CI = .23, .58) and baseline MST performance ($\beta = .64$; 95% CI = .49, .80), but not mental toughness ($\beta = .08$; 95% CI = -.08, .24), were credible predictors and accounted for 63% of the variance in MST performance when fatigued. The correlation between $\dot{V}O_{2\text{peak}}$ and mental toughness was inverse and small ($r = -.04$; 95% CI = -.38, .31); $\dot{V}O_{2\text{peak}}$ and baseline MST was moderate-to-large ($r = .50$; 95% CI = .20, .72); and between baseline MST and mental toughness was small ($r = .13$; 95% CI = -.23, .45).

Consistent with recommendations for Bayesian analysis^{18,19}, we conducted a series of analyses to examine the sensitivity of the informative priors on the posterior distribution. For RQ1, we first examined the influence of different values for the prior variance using a well-specified prior mean for the regression paths of $\dot{V}O_{2\text{peak}}$ ($\mu_1 = .50$) and mental toughness ($\mu_2 = .15$) that was equivalent to the robust maximum likelihood (MLR) estimator, thus ensuring complete congruence with the data. As depicted in Figure 1a, greater certainty in the prior (i.e., smaller variance) is associated with enhanced precision in the posterior distribution. Second, we examined the relative stability of the results by varying the mean and/or variance of the effect of mental toughness on MST performance. With regard to RQ1, there were minimal differences in the posterior mean effect and that of the MLR estimate ($\mu = .15$) with

a prior variance of 1 (see Figure 1b). However, differences between the MLR and Bayesian estimates occurred for all other variations in the mean and variance of the effect of mental toughness on MST performance. When $\mu_2 = .50$, the effect was credible when the prior variance was less than .10, whereas the effect was significant with $\mu_2 = .15$ only for those priors with high precision ($\sigma^2 = .005$; $\sigma^2 = .001$). Overall, the sensitivity analyses indicated that the prior specification of μ_2 had more influence on the posterior results when there was greater certainty in those beliefs in terms of the strength, precision (i.e., credibility intervals), and significance of the effect. Similar findings were observed for RQ2; notably, there are minimal differences in the posterior mean effect and that of the MLR estimate ($\mu = -.07$) with a prior variance of 1 (see Figure 1b). The greatest stability in the posterior distribution was observed with a prior mean of zero for the effect of mental toughness on behavioural perseverance.

Discussion

The purpose of this study was to conduct a conceptual replication of the association between mental toughness and behavioural perseverance, and test the robustness of this effect when athletes were physically fatigued. First, we provided strong evidence of an association between mental toughness and behavioural perseverance under typical circumstances (i.e., little or no unique challenge to performance execution), when prior knowledge from past research is incorporated directly into the estimation process. Second, our findings suggest that the association between mental toughness and behavioural perseverance does not generalise to a performance context in which the MST was employed as a test of the colloquial “fourth quarter” effort in a football match (i.e., fatiguing protocol simulated the physical demands of approximately $\frac{3}{4}$ of game).

Past work has underscored the importance of mental toughness for behavioural perseverance across a range of achievement settings, including students’ academic and social

goal progress across a university semester⁵, soldiers' performance in a 6-week selection test in the military⁵, and athletes' scores on a discrete tasks that requires them to 'push through' physical fatigue^{3,4}. In contrast to this body of work, the results of the current study offered equivocal evidence for the notion that mental toughness enhances the likelihood of behavioural perseverance. Specifically, we found evidence for this theoretical proposition under typical performance circumstances, but not so when footballers were physically fatigued. With regard to baseline behavioural perseverance prior to fatiguing, the current findings can be considered an update on existing knowledge using new data to produce results that are akin to an automatic meta-analysis²⁰. The robustness of the association when prior knowledge was incorporated directly into the estimation process to different forms of prior information indicated some degree of instability, such that the greatest stability in the findings was evident when the standardised mean centred on .15 (relative to the prior expectation of .24). This finding is unsurprising given that we controlled for footballers' aerobic capacity in the analysis. As such, these results can be considered the most up-to-date knowledge regarding the size of association between mental toughness and behavioural perseverance.

Despite the encouraging findings for RQ1, our results raise questions regarding the robustness of the association between mental toughness and behavioural perseverance when the 'going gets tough'^{10,21,22}. Although the effect was in the expected direction (i.e., higher mental toughness associated with better MST performance), the size of the association was small and untrustworthy. Sensitivity analyses indicated that the most stable effect hovered around zero. These findings could be an indication of the instability of this theoretical feature of mental toughness, or the methodological procedures employed in this study. First, the conceptualisation of mental toughness as a psychological capacity of individuals that characterises their potential for action towards an objective implies that people are aware of

the need to activate this resource and how it can be applied effectively within the confines of the situational demands. Such an interpretation is consistent with a state-like conceptualisation in that mental toughness has some enduring properties across contexts and time yet varies depending on the features of the situation⁵. Second, given the goal-directed nature of mental toughness, it may be that the footballers' performance when fatigued was incompatible with their self-referenced goal, relative to their baseline performance. In other words, participants may have persevered to an extent that was consistent with their goal, yet their goal attainment was captured inaccurately in the absolute measure of MST performance (shuttle:level).

There are several ways by which the limitations of our work can be considered in future research. First, manipulating participants' awareness of situational demands and the salience of mental toughness for task performance will permit a direct test of a core conceptual feature (i.e., psychological capacity or potential for action). Relatedly, assessing participants' mental toughness prior to each testing session would provide insight regarding the temporal dynamics of the association between mental toughness and perseverance (e.g., do changes in rather than the level of mental toughness drive the effect?). Second, measuring prospectively participants' goal for their performance in the MST under typical circumstances and when fatigued, relative to their underlying aerobic capacity, will provide an individualised assessment of performance that is relative to the task. Third, as mental toughness is conceptualised as potential versus realised action, it is important to consider variables (e.g., motivation) and experimental protocol enhancements (e.g., increased investment in or importance of the task²⁴) that might maximise the mobilisation of this psychological capacity for performance.

Conclusions

Persistence and perseverance towards important performance goals is an attribute of people that is much coveted by those who struggle with their goal pursuits. Broadly, speaking, research supports the importance of psychological skills for sport performance²⁵ and, in particular, endurance sports that require perseverance towards task goals²⁶.

Representing one such psychological resource, the results of the current study suggest that mental toughness is salient for behavioural perseverance in a discrete physical task that taxes the aerobic energy system. However, the findings also shed light on a potential boundary condition to this theoretical proposition – who, where, when factors²⁷ – that requires additional testing in future research.

Practical Implications

- Underlying fitness ($\dot{V}O_{2\text{peak}}$) is important for behavioural perseverance under typical circumstances and when fatigued
- Mental toughness is important for behavioural perseverance during typical performance settings but not when fatigued
- The STGC is a feasible experimental protocol for fatiguing participants within the context of Australian rules football

Acknowledgements

The corresponding author is supported by a Curtin Research Fellowship.

References

1. Gucciardi DF. Mental toughness: progress and prospects. *Current Opin Psychol* 2017; 17-23. doi: 10.1016/j.copsyc.2017.03.010
2. Cowden RG. Mental toughness and success in sport: a review and prospect. *Open Sports Sci J* 2017; 10:1-14.
3. Bell JJ, Hardy L, Beattie S. Enhancing mental toughness and performance under pressure in elite young cricketers: a 2-year longitudinal intervention. *Sport, Exercise Perform Psychol* 2014; 2(2): 281-297. doi: 10.1037/a0033129
4. Gucciardi DF, Peeling P, Ducker K, et al. When the going gets tough: Mental toughness and its relationship with behavioural perseverance. *J Sci Med Sport* 2016; 19(1): 81-86. doi: 10.1016/j.jsams.2014.12.005
5. Gucciardi DF, Hanton S, Gordon S, et al. The concept of mental toughness: tests of dimensionality, nomological network, and traitness. *J Pers* 2015;83(1);26-44.
6. Morwitz VG, Fitzsimons GJ. The mere-measurement effect: why does measuring intentions change actual behaviour. *J Consum Psychol* 2004;14(1&2):64-73.
7. Drust B, Waterhouse G, Atkinson G, et al. Circadian rhythms in sports performance – an update. *Chronobiol Int* 2005; 22(1): 21-44.
8. Cook C, Crust L, Littlewood M, et al. ‘What it takes’: perceptions of mental toughness and its development in an English Premier league soccer academy. *Qual Res Sport Exerc Health* 2014; 6(3): 329-347. doi: 10.1080/2159676X.2013.857708
9. Jaeschke A-MC, Sachs ML, Dieffenbach KD. Ultramarathon runners’ perceptions of mental toughness: a qualitative inquiry. *Sport Psychol* 2016; 30(3): 242-255. doi: 10.1123/tsp.2014-0153

10. Jones G, Hanton S, Connaughton D. What is this thing called mental toughness? an investigation of elite sport performers. *J App Sport Psychol* 2002; 14(3): 205-218. doi: 10.1080/10413200290103509
11. Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. *Struct Equ Modeling* 2002;9(4):599 – 620. doi: 10.1207/S15328007SEM0904_8
12. Jolley D, Dawson B, Maloney SK, et al. Hydration and urinary pseudoephedrine levels after a simulated team game. *Int J Sport Nutr Exerc Metab* 2014;24(3):325-333.
13. Muthén LK, Muthén B. *Mplus users guide* (7th ed.). Los Angeles CA: Muthén & Muthén. 2012.
14. Muthén BO, Asparouhov T. Bayesian structural equation modelling: a more flexible representation of substantive theory. *Psychol Methods* 2012; 17(3): 313-335.
15. Gelman A, Carlin JB, Stern HS et al. *Bayesian data analysis*, Florida, Chapman and Hall/CRC. 2013.
16. Etz A, Vandekerckhove J. Introduction to Bayesian inference for psychology. *Psychon Bull Rev* in press. doi: 10.3758/s13423-017-1262-3
17. Kruschke JK, Liddell TM. Bayesian data analysis for newcomers. *Psychon Bull Rev* in press. doi: 10.3758/s13423-017-1272-1
18. Depaoli S, van de Schoot R. Improving transparency and replication in Bayesian statistics: the WAMBS-checklist. *Psychol Methods* in press. doi: 10.1037/met0000065
19. van de Schoot R, Broere JJ, Perryck KH, et al. Analysing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical

- ventilation in burn survivors. *Eur J Psychotraumatol* 2015; 6: 25216. doi: 10.3402/ejpt.v6.25216
20. Zyphur MJ, Oswald FL. Bayesian estimation and inference: A user's guide. *J Manage* 2015; 41(2): 390-420. doi: 10.1177/0149206313501200
21. Gucciardi DF, Gordon S, Dimmock JA. Towards an understanding of mental toughness in Australian football. *J App Sport Psychol* 2008; 20(3): 261-281.
22. Weinberg R, Butt J, Culp, B. Coaches' views of mental toughness and how it is built. *Int J Sport Exercise Psychol* 2011; 9(2): 156-172. doi: 10.1080/1612197X.2011.567106
23. Mayorga-Vega D, Aguilar-Soto P, Viciano J. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: a meta-analysis. *J Sports Sci Med* 2015;14(3):536-547.
24. Mesagno C, Harvey JT, Janelle CM. Self-presentation origins of choking: evidence from separate pressure manipulations. *J Sport Exerc Psychol* 2011; 33(3):441-459.
25. Brown DJ, Fletcher D. Effects of psychological and psychosocial interventions on sport performance: a meta-analysis. *Sports Med* 2017;47(1):77-99.
26. McCormick A, Meijen C, Marcora S. Psychological determinants of whole-body endurance performance. *Sports Med* 2015;45(7):997-1015.
27. Busse C, Kach AP, Wagner SM. Boundary conditions: what they are, how to explore them, why we need them, and when to consider them. *Organ Res Methods* in press. doi: 10.1177/1094428116641191

Table 1. Overview of priors employed in Bayesian analyses of a priori models.

Parameter	RQ1 (baseline MST)			
	μ	σ^2	Form	Type
MST (μ)	0	10^{10}	N	D
$\dot{V}O_{2\text{peak}}$ (μ)	0	10^{10}	N	D
MT (μ)	0	10^{10}	N	D
$\dot{V}O_{2\text{peak}} \rightarrow$ MST (β)	.60	.01	N	E ²³
MT \rightarrow MST (β)	.24	.005	N	E ²
MST (σ^2)	-1	0	IG	D
MT (σ^2)	0	-3	IW	D
$\dot{V}O_{2\text{peak}}$ (σ^2)	0	-3	IW	D
$\dot{V}O_{2\text{peak}} \leftrightarrow$ MT	0	-3	IW	D
RQ2 (MST when fatigued)				
	μ	σ^2	Form	Type
MST _{fatigued} (μ)	0	10^{10}	N	D
$\dot{V}O_{2\text{peak}}$ (μ)	0	10^{10}	N	D
MST _{baseline} (μ)	0	10^{10}	N	D
MT (μ)	0	10^{10}	N	D
$\dot{V}O_{2\text{peak}} \rightarrow$ MST _{fatigued} (β)	.60	.01	N	E ²³
MST _{fresh} \rightarrow MST _{fatigued} (β)	.70	.01	N	E ²
MT \rightarrow MST _{fatigued} (β)	.15	.01	N	E
MST _{fatigued} (σ^2)	-1	0	IG	D
$\dot{V}O_{2\text{peak}}$ (σ^2)	0	-4	IW	D
MST _{baseline} (σ^2)	0	-4	IW	D
MT (σ^2)	0	-4	IW	D
$\dot{V}O_{2\text{peak}} \leftrightarrow$ MST _{baseline}	0	-4	IW	D
$\dot{V}O_{2\text{peak}} \leftrightarrow$ MT	0	-4	IW	D
MST _{baseline} \leftrightarrow MT	0	-4	IW	D

Note: μ = mean; σ^2 = variance; Form = distributional form of the prior; IG = Inverse Gamma; IW = Inverse Wishart; E = empirically-informed prior (informative); D = default prior (uninformative).

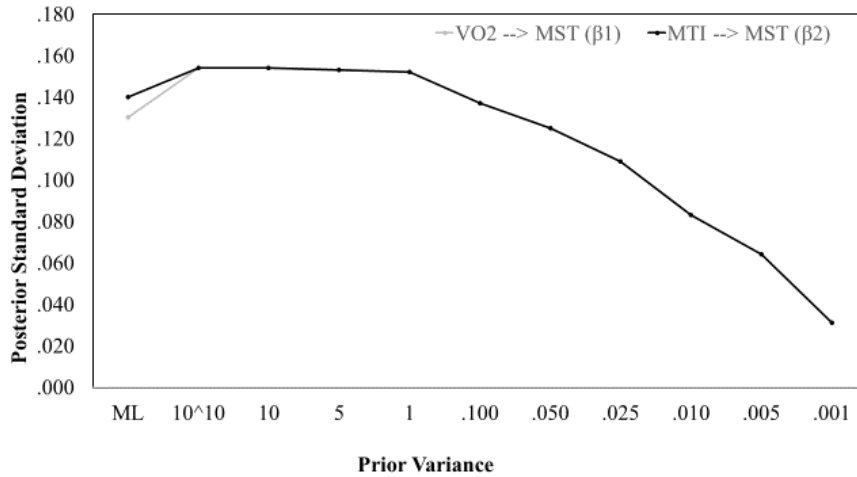


Figure 1a: Effect of different estimators (MLR vs Bayes) and different variances, but with a fixed mean for β_1 and β_2 (based on MLR output) for RQ1. Note: MLR = robust maximum likelihood estimator.

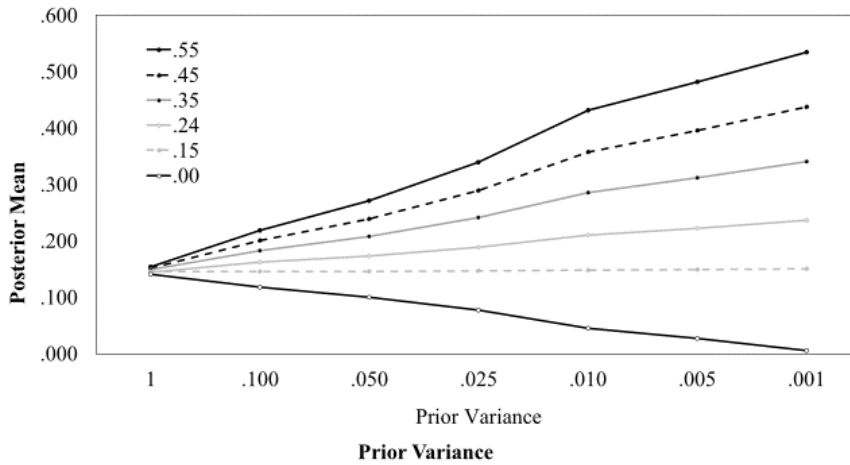


Figure 1b: Effect of different variances on the prior mean value of mental toughness on 20-m multistage shuttle run test performance (RQ1).

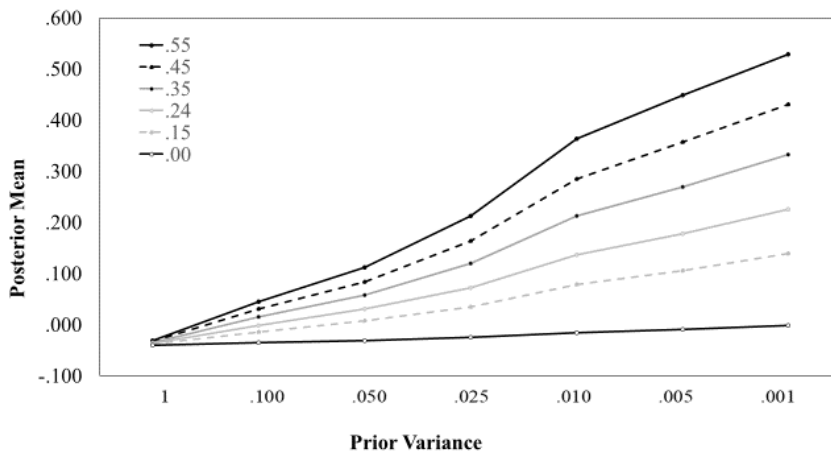


Figure 1c: Effect of different variances on the prior mean value of mental toughness on 20-m multistage shuttle run test performance (RQ2).

Supplementary Material

Section 1 – Monte Carlo Simulations for Sample Size Planning

Guided by recommendations for power calculations and sample size planning¹, we conducted a series of Monte Carlo simulations using maximum likelihood and Bayesian estimators (i.e., 1000 replications, variables standardised with a mean of 0 and variance of 1). Simulations with a robust maximum likelihood estimator (MLR) indicated that 35 participants would provide 99% power detect the expected effect of $\dot{V}O_{2\text{peak}}$ on MST performance ($\beta = .60$) (i.e., is significant in 99% of the simulated datasets), where the 95% coverage of the population value across replications exceeds the recommended value of .90¹; however, this sample size would provide only 52% power for the expected effect of mental toughness on MST performance ($\beta = .24$), yet the 95% coverage was sufficient (i.e., 90%). A sample size of 71 participants would provide 80% power detect an effect of $\beta = .24$ for mental toughness on MST performance (95% coverage = .93).

We replicated these simulations with a Bayesian estimator using a fixed number of 1000 iterations across each of the four Markov Chain Monte Carlo (MCMC) chains. First, as a comparison with the MLR results detailed above, we started with a sample size of 35 and employed default, non-informative priors for all parameters in the model. This model provided 99% power with .95 for the 95% coverage of the population to detect the expected effect of $\dot{V}O_{2\text{peak}}$ on MST performance ($\beta = .60$), with only 42% power and .95 for the 95% coverage of the population value to detect the expected effect ($\beta = .24$). Using non-informative priors on all parameter estimates, a sample size of 78 participants would provide 80% power detect an effect of $\beta = .24$ for mental toughness on MST performance (95% coverage = .95). These findings are consistent with related simulation work in which it was shown that both maximum likelihood and Bayesian estimation exhibit lower power with small samples². It is therefore the incorporation of prior information directly into the

estimation that enables researchers to capitalise on the strengths of Bayesian estimation³. As this study built on existing work, we had access to prior information in the form of empirical results for the effects of $\dot{V}O_{2\text{peak}}$ ($\mu = .60, \sigma^2 = .01$)⁴ and mental toughness on MST performance ($\mu = .24, \sigma^2 = .005$)⁵. Incorporating these informative priors directly within the Monte Carlo simulations indicated that 6 participants would provide 100% power with 95% coverage of the population value (1.00) to detect these a priori effects. Subsequently, we increased the prior variance for each parameter threefold to model greater uncertainty in these prior estimates of the mean value ($\dot{V}O_{2\text{peak}}$ $\sigma^2 = .03$; mental toughness $\sigma^2 = .015$); this simulation showed that 27 participants would provide sufficient power (81%) and 95% coverage of the population value (1.00) to detect the expected effect of mental toughness on MST performance. Based on these simulations, we decided to recruit into this study at least 30 footballers. The *Mplus* syntax is provided below in Tables S1 and S2.

Table S1. *Mplus syntax for Monte Carlo simulations with a robust maximum likelihood estimator.*

TITLE: Monte Carlo simulation for planning sample size (MLR estimator)

MONTECARLO:

NAMES = VO2 MTlevel BT1_Dec;
NOBSERVATIONS = 71;
SEED = 500;
NREPS = 1000;

MODEL POPULATION:

[VO2*0 MTlevel*0];
VO2*1 MTlevel*1;
[BT1_Dec*0];
BT1_Dec*.50;
BT1_Dec ON VO2*.60;
BT1_Dec ON MTlevel*.24;
MTlevel WITH VO2*.15;

MODEL:

[BT1_Dec*0];
BT1_Dec*.50;
BT1_Dec ON VO2*.60 (b1);
BT1_Dec ON MTlevel*.24 (b2);
MTlevel WITH VO2*.15;

ANALYSIS:

ESTIMATOR = MLR;

OUTPUT: tech9;

Table S2. *Mplus syntax for Monte Carlo simulations with a Bayesian estimator (Note: text in green and preceded by an exclamation mark is not read by Mplus when executing the analysis).*

TITLE: Monte Carlo simulation for planning sample size (Bayes estimator)

MONTECARLO:

NAMES = VO2 MTlevel BT1_Dec;
 NOBSERVATIONS = 27;
 SEED = 500;
 NREPS = 1000;

MODEL POPULATION:

[VO2*0 MTlevel*0];
 VO2*1 MTlevel*1;
 [BT1_Dec*0];
 BT1_Dec*.50;
 BT1_Dec ON VO2*.60;
 BT1_Dec ON MTlevel*.24;
 MTlevel WITH VO2*.15;

MODEL:

[BT1_Dec*0];
 BT1_Dec*.50;
 BT1_Dec ON VO2*.60 (b1);
 BT1_Dec ON MTlevel*.24 (b2);
 MTlevel WITH VO2*.15;

ANALYSIS:

ESTIMATOR = BAYES;
 FBITERATIONS = 1000;
 PROCESSORS = 4;
 CHAINS = 4;

MODEL PRIORS:

! informative prior for b1 (<https://www.ncbi.nlm.nih.gov/pubmed/26336340>)

b1~N(.60,.03);

! informative priors for b2 (doi: 10.1016/j.jsams.2014.12.005)

b2~N(.24,.015);

OUTPUT: tech9;

Section 2 – Details of the Individual Components of the Study Protocol

Graded Exercise Test (GXT). The GXT was performed on a motorised treadmill (VR3000, NuryTech Inc., Japan) using 3 min exercise and 1 min rest periods. The initial speed was set at 12 km·h⁻¹ with subsequent increments of 2 km·h⁻¹ for the first two stages, and 1 km·h⁻¹ thereafter until voluntary exhaustion. Throughout the GXT, expired air was analysed for concentrations of O₂ and CO₂ (Ametek Gas Analysers, Applied Electrochemistry, SOVS-3A/1 and COV CD-3A, Pittsburgh, PA, USA). The gas analysers were calibrated pre-test and verified post-test with certified gravimetric gas mixtures (BOC Gases, Chatswood, Australia). Ventilation was recorded at 15 s intervals via a turbine ventilometer (Vacumed Universal Ventilation Meter, 17125, Ventura, CA), which was also calibrated before, and verified after exercise using a 1 L syringe. Determination of peak oxygen consumption ($\dot{V}O_{2peak}$) was attained by summing the four highest consecutive 15 $\dot{V}O_2$ values obtained throughout the test. Additionally, measures of blood lactate (BLa), heart rate (HR) and a rating of perceived exertion (RPE) were collected during each 1 min recovery period throughout the GXT.

Multi-stage 20m Shuttle Run (MST). Prior to completing the MST, participants were guided through a standardised 5 min warm-up, consisting of 2 min of low intensity jogging, 2 min of 20 m dynamic exercises and 1 min of high intensity, football specific, dynamic movements. The MST involved repeated sprint efforts where participants ‘shuttled’ between a set of lines 20 m apart to the pace of a “beep” on a pre-recorded audio track⁶ (Brewer et al., 1988). Upon failure to complete two consecutive shuttles, the test was concluded and measures of HR, RPE and BLa were collected.

Simulated Team Game Circuit (STGC). The STGC was designed to simulate the demands of Australian football in accordance with the protocol of Jolley et al. (2014)⁷. Prior

to completing the STGC, participants were guided through a standardised dynamic warm-up (as outlined above). The STGC required participants to complete 3 sets of 20 x 92 m circuits. Each circuit took approximately 35 s to complete, and involved intermittent periods of walking, jogging, striding, sprinting, agility, jumping, bumping and tackling (see Figure S1). A new circuit commenced every 60 s, leaving approximately 25 s of rest between each circuit. A 5 min rest period was allowed after the first and third set, and a 10 min rest period was allowed after set two. Following the 5 min break permitted after set three of the STGC, participants were required to complete a 20 m MST. Since the 3 x 20 min sets, along with their subsequent rest periods, were designed to simulate three quarters of a football game; the MST was used as a quantifiable test of the colloquial “fourth quarter” effort. Throughout the STGC, HR and RPE were recorded at the completion of each set, and BLa was collected at the completion of set 3, and again after the 20 m MST.

Blood Sampling and Analysis. Blood samples (5 μ L) were drawn from the participants’ ear lobe to assess blood lactate (BLa). Initially, the collection site was wiped clean with an alcohol swab before a small incision was made (pin-prick) using a spring loaded lancet. The first drop of blood produced was wiped away, and the second drop was collected into a lactate test meter (Lactate Pro 2; Arkray, Japan).

Heart Rate and Perceived Exertion. Participants’ HR was measured using a Garmin heart rate monitor (Forerunner 610, Switzerland), and rate of perceived exertion (RPE) was recorded via the Borg 6-20 perceptual scale⁸. This scale encompassed the anchor points of 6 (very, very light) and 20 (maximal exertion).

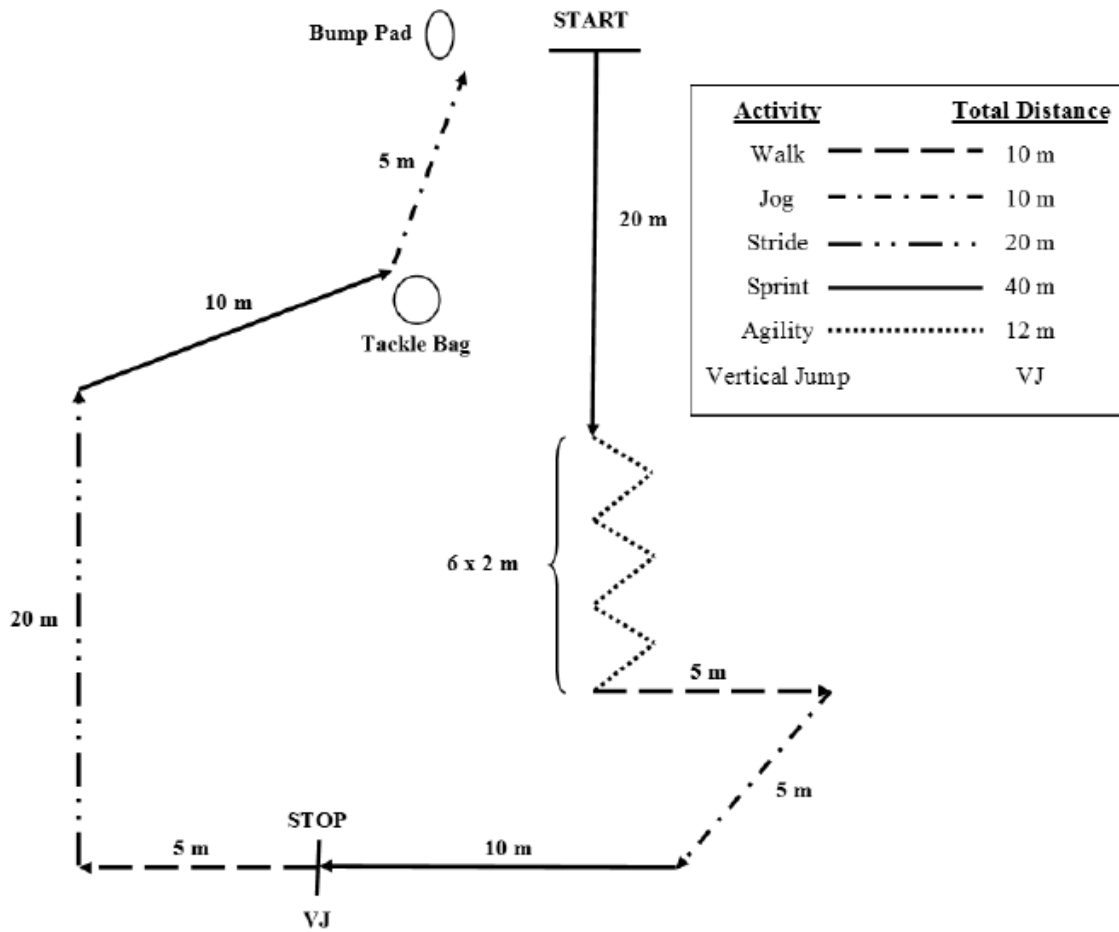


Figure S1. Simulated team game circuit (adapted from Jolley et al., 2014). Note: not to scale.

Section 3 – Checks of Testing Protocol

A key assumption of our testing protocol is the validity of the STGC as a protocol for fatiguing participants. To check the veracity of this assumption, we used latent growth modelling with a Bayesian estimator in *Mplus* 8 to determine the amount of individual change between the two time testing sessions with regard to MST performance, rate of perceived exertion, BLa, and heart rate⁹. Default, non-informative priors were used so as to allow the data to drive the estimation process.

Validity of the STGC Protocol

The probability of the hypothesised model, given the data, was acceptable (PPP = .447, Δ observed and replicated χ^2 95% CI [-9.28, 11.11]). Footballers' performance on the MST when fresh is captured by the intercept factor ($\tau = 12.48$ [95% CI = -1.65, -1.03]). The mean of the latent slope factor ($\tau = -1.342$ [95% CI = .42, 2.01]) is equivalent to the overall intra-individual mean difference between testing sessions, thus indicating that footballers' performance on the MST decreased on average.

Effort Expended during Testing Sessions

As each of the tests executed in this study depend on the extent to which sufficient effort was expended by participants during each activity, we conducted a series of analyses to assess the degree to which footballers pushed towards exhaustion. Collectively, these findings provided psychological and physiological support to indicate that participants interpreted a high degree of sensory input from various physiological systems and exerted sufficient effort to regulate these markers via exercise intensity production.

Perceived Exertion. Participants' reports of their RPE ranged from 15 to 20 (19.05 ± 1.25) for the lab-based GXT; 14 to 19 (17.87 ± 1.40) for the MST when fresh; and 15 to 20 (18.63 ± 1.12) for the MST when fatigued. With regard to a comparison of MST performance under typical conditions and when fatigued, the probability of the hypothesised model, given

the data, was acceptable (PPP = .445, Δ observed and replicated χ^2 95% CI [-9.25, 11.06]). The intercept ($\tau = 17.87$ [95% CI = 17.38, 18.35]) and slope factors ($\tau = .76$ [95% CI = .27, 1.26]) indicated that participants, on average, reported a higher level of exerted effort in the MST when fatigued.

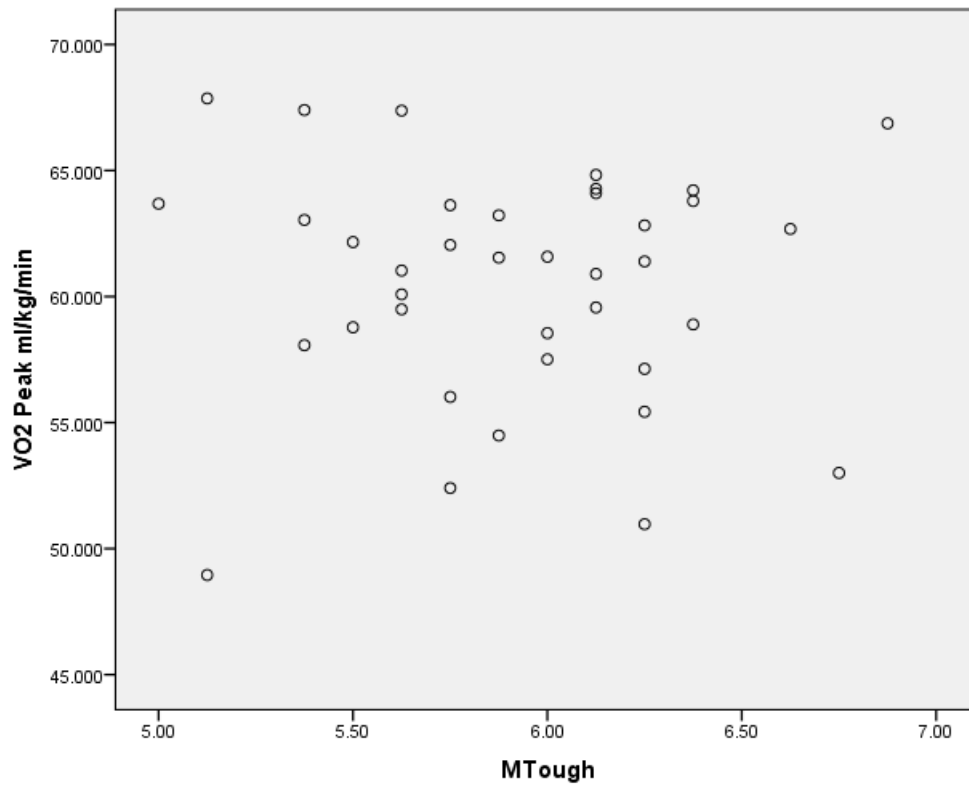
BLa. Blood analyses indicated that BLa ranged from 5.3 to 16.2 mmol (10.41 ± 2.92) for the lab-based GXT; 4.6 to 15.7 mmol (8.62 ± 2.85) for the MST when fresh; and 2.2 to 13.2 mmol (6.20 ± 2.32) for the MST when fatigued. With regard to a comparison of MST performance under typical conditions and when fatigued, the probability of the hypothesised model, given the data, was acceptable (PPP = .445, Δ observed and replicated χ^2 95% CI [-9.26, 11.02]). The intercept ($\tau = 8.62$ [95% CI = 7.62, 9.62]) and slope factors ($\tau = -2.43$ [95% CI = -3.53, -1.31]) indicated that participants' BLa, on average, decreased between the two sessions. Guided by recommendations for blood lactate diagnostics in exercise testing¹⁰, we considered BLa > 4 as indicative of an effort in response to the incremental intensity of the test. These findings indicated that sufficient effort was expended by participants across all trials, with the exception of 4 participants with BLa levels of 2.2 to 3.9 mmol in the MST when fatigued. The exclusion of these participants from the main analyses did not alter the strength or significance of effects.

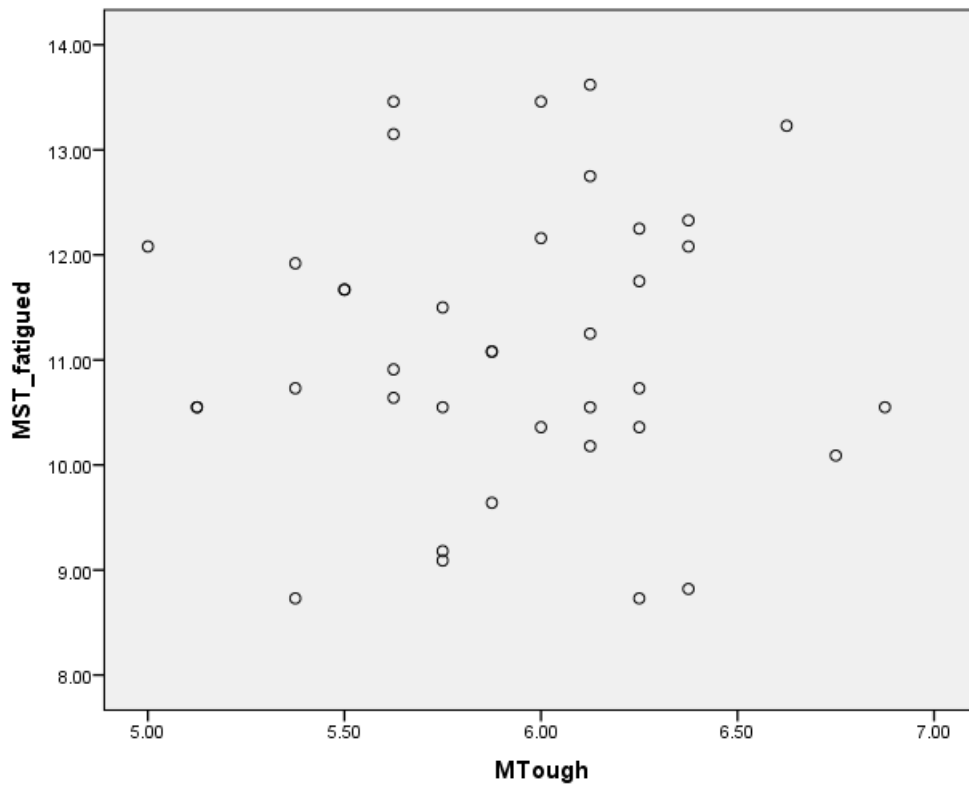
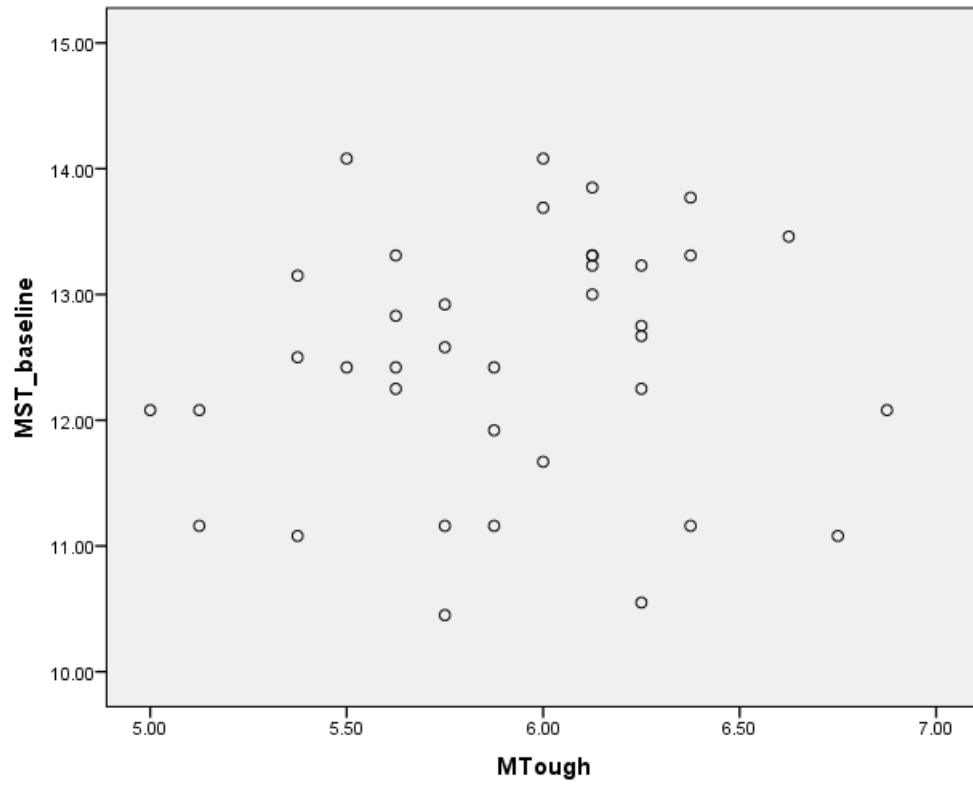
Heart Rate. Raw maximum heart rate scores ranged from 176 to 211 (192.95 ± 7.87) for the lab-based GXT; 176 to 211 (191.08 ± 8.68) for the MST when fresh; and 163 to 209 (188.89 ± 8.90) for the MST when fatigued. With regard to a comparison of MST performance under typical conditions and when fatigued, the probability of the hypothesised model, given the data, was acceptable (PPP = .445, Δ observed and replicated χ^2 95% CI [-9.26, 11.01]). The intercept ($\tau = 191.08$ [95% CI = 188.03, 194.11]) and slope factors ($\tau = -2.19$ [95% CI = -5.20, .087]) indicated that participants' heart rate, on average, was slightly lower in the MST session when fatigued though this difference was untrustworthy. In total,

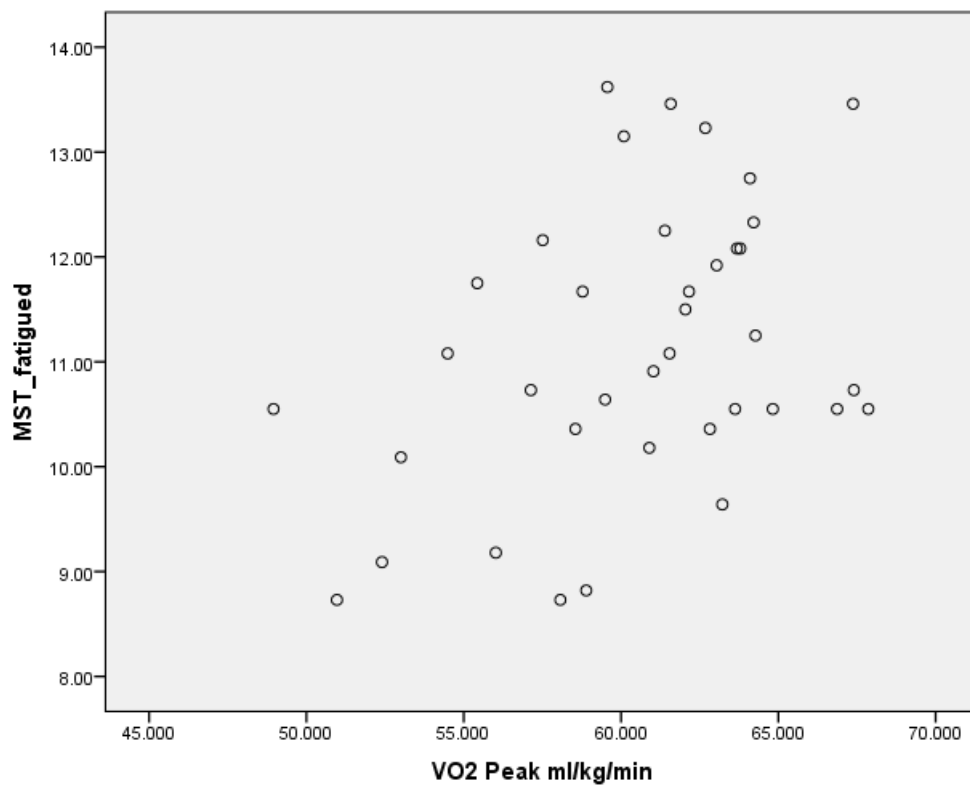
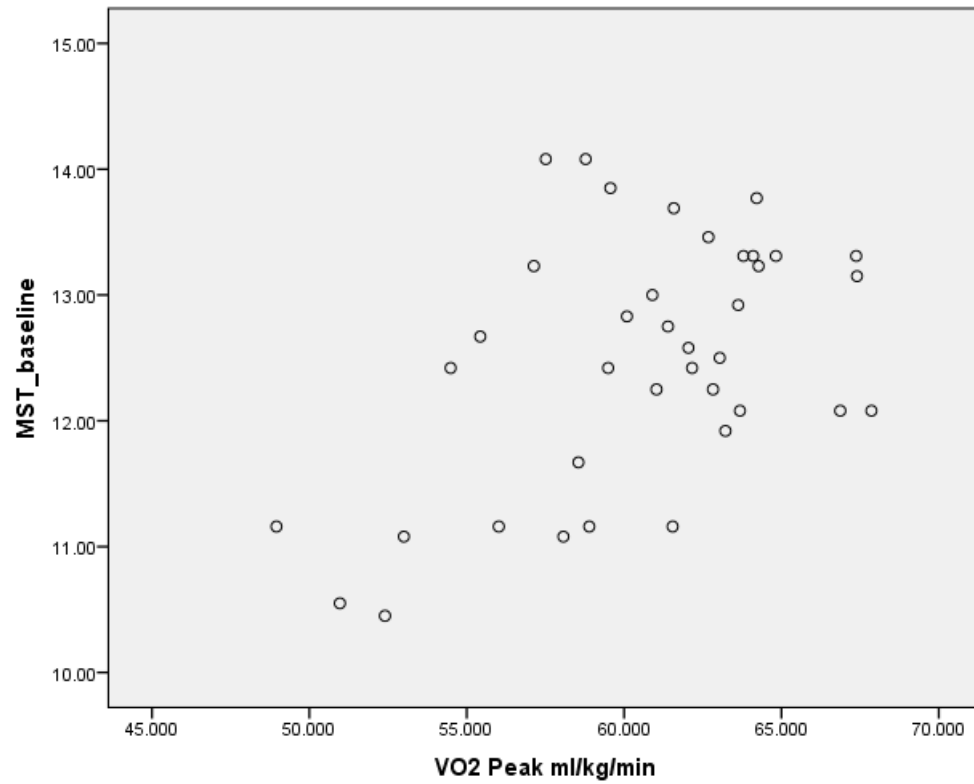
20 participants reached or exceeded a maximal heart rate that was higher than their $HR_{max_{pred}} (208 - [0.7 * Age])^{11}$ during the GXT; 14 footballers during the MST when fresh; and 11 players during the MST when fatigued. Of those footballers who did not reach or exceed their $HR_{max_{pred}}$, the relative percentage attained was greater than 92% in the GXT, 91% in the MST when fresh, and 84% in the MST when fatigued.

Section 4 – Plots of Raw Data and Posterior Distributions

At the suggestion of a reviewer, we included plots of (i) raw data among the study variables, and (ii) prior and posterior densities of the direct effect parameters for both research questions. These figures are displayed in the following sections.







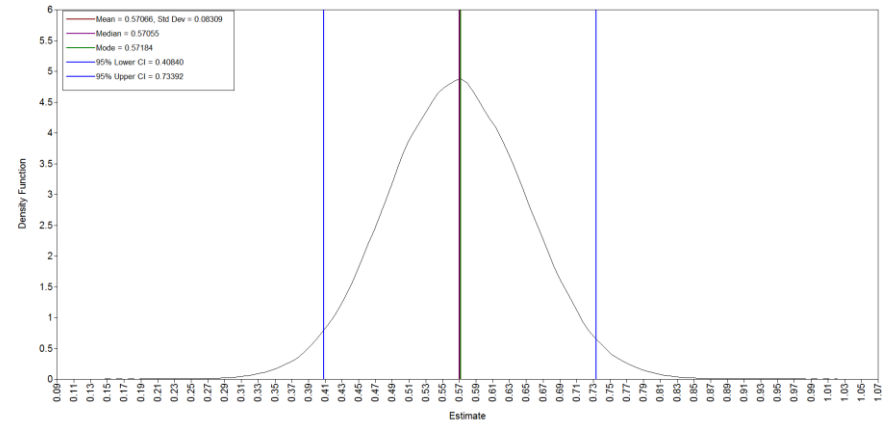
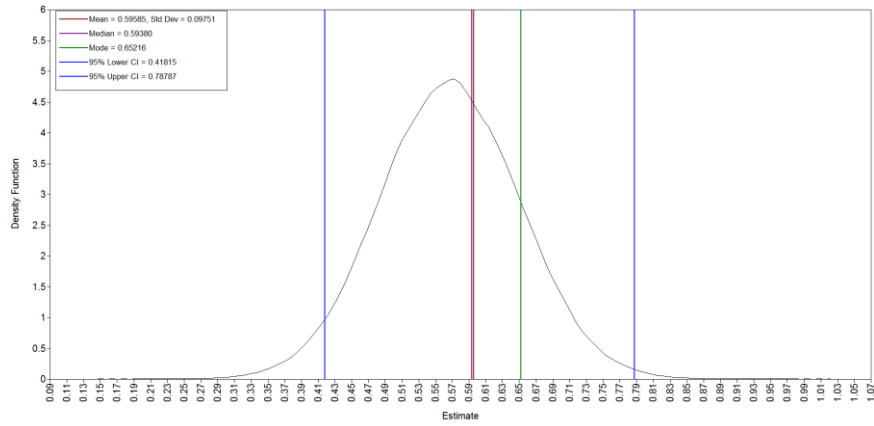


Figure S2a. Prior (left-hand side) and posterior (right-hand side) density distributions of the direct effect of baseline MST on VO₂.

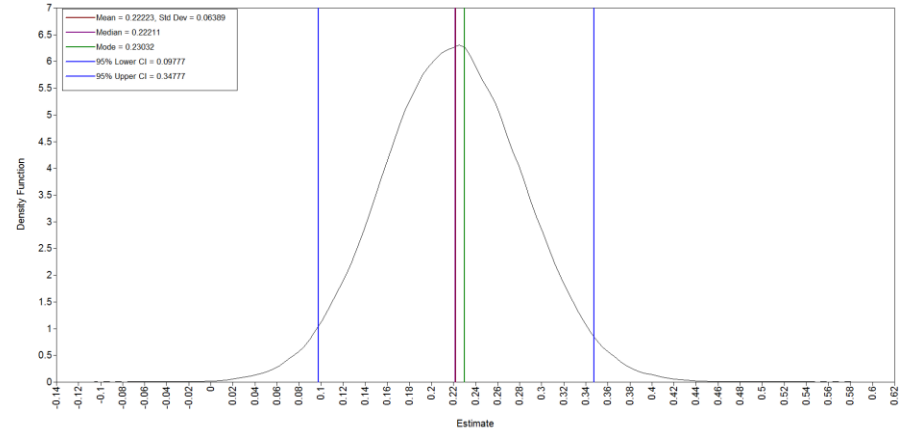
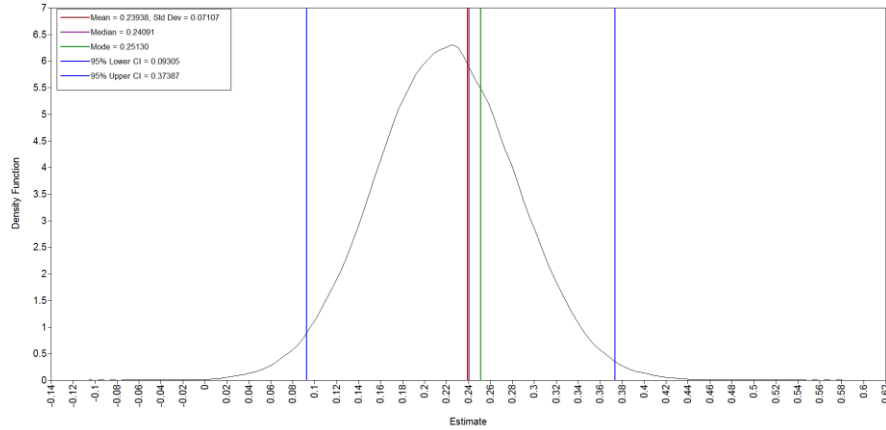


Figure S2b. Prior (left-hand side) and posterior (right-hand side) density distributions of the direct effect of baseline MST on mental toughness.

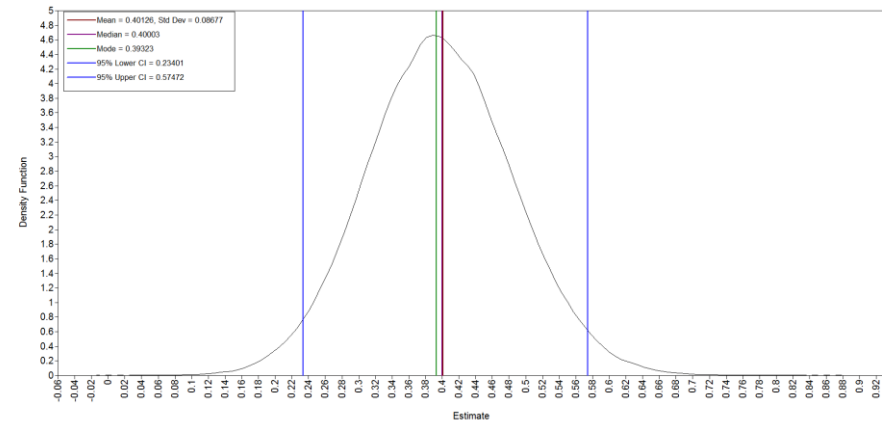
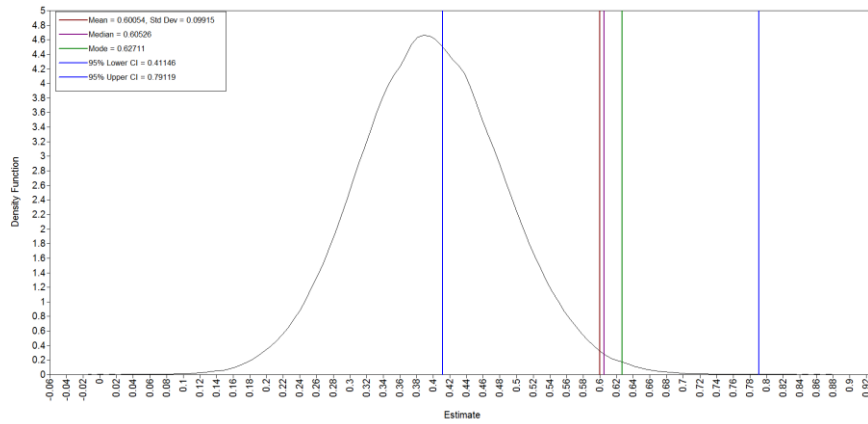


Figure S3a. Prior (left-hand side) and posterior (right-hand side) density distributions of the direct effect of MST when fatigued on VO2.

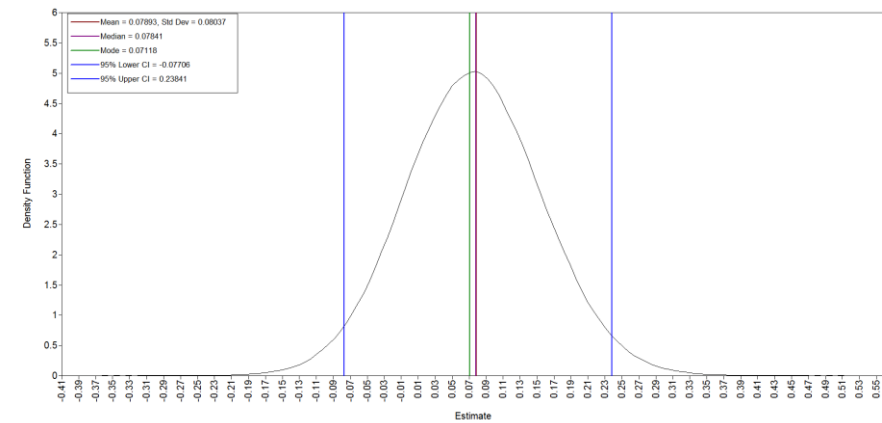
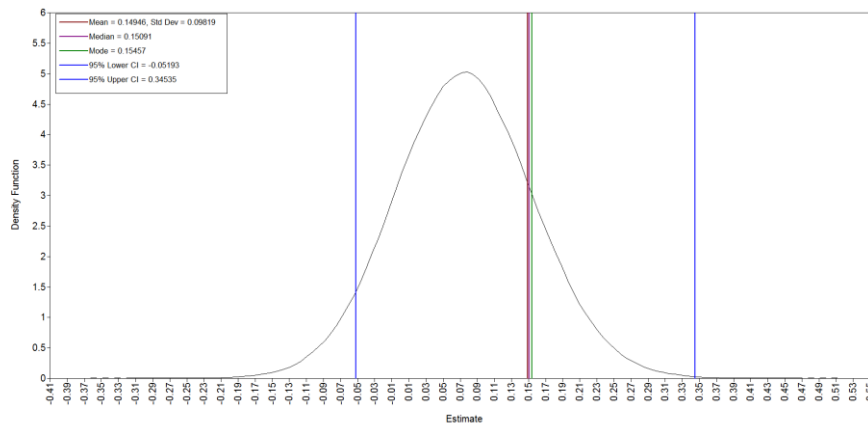


Figure S3b. Prior (left-hand side) and posterior (right-hand side) density distributions of the direct effect of MST when fatigued on mental toughness.

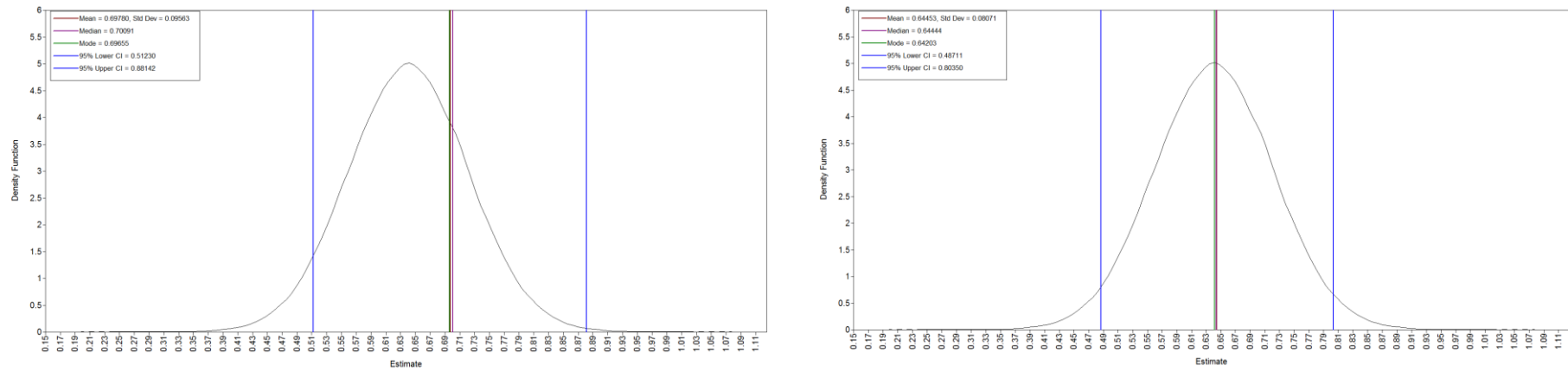


Figure S3c. Prior (left-hand side) and posterior (right-hand side) density distributions of the direct effect of MST when fatigued on baseline MST.

References

1. Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. *Struct Equ Modeling* 2002;9(4):599 – 620. doi: 10.1207/S15328007SEM0904_8
2. van de Schoot R, Broere JJ, Perryck KH, et al. Analysing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *Eur J Psychotraumatol* 2015; 6: 25216. doi: 10.3402/ejpt.v6.25216
3. McNeish D. On using Bayesian methods to address small sample problems. *Struct Equ Modeling* 2016;23(5):750-773.
4. Mayorga –Vega D, Aguilar-Soto P, Viciano J. Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: a meta-analysis. *J Sports Sci Med* 2015;14(3):536-547.
5. Gucciardi DF, Peeling P, Ducker K, et al. When the going gets tough: Mental toughness and its relationship with behavioural perseverance. *J Sci Med Sport* 2016; 19(1): 81-86. doi: 10.1016/j.jsams.2014.12.005
6. Brewer J, Ramsbottom R, Williams C. *Multistage Fitness Test*, Loughborough, National Coaching Foundation; 1988.
7. Jolley D, Dawson B, Maloney SK, White J, Goodman C, Peeling P. Hydration and urinary pseudoephedrine levels after a simulated team game. *Int J Sport Nutr Exerc Metab* 2014;24(3):325-333.
8. Borg G. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc* 1982; 14(5):377-381.
9. Voelkle MC. Latent growth curve modeling as an integrative approach to analysis of change. *Psychol Sci* 2007;49(4):375-414.

10. Beneke R, Leithauser RM, Ochentel O: Blood lactate diagnostics in exercise testing and training. *Int J Sports Physiol Perform* 2011;6(1): 8-24.
11. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. *J Am Coll Cardiol* 2001;37(1):153–156. doi:10.1016/S0735-1097(00)01054-8