SportRXiv

Preprint
Not Peer Reviewed

Capturing the 'expert's eye': Towards a better understanding and implementation of subjective performance evaluations in team sports

Johann Windt [1,2], Keith Hamilton[3], David N. Cox[1,3], Bruno D. Zumbo[4], Ben Sporer[1,5]

[1]*Vancouver Whitecaps Football Club, Vancouver, Canada*
[2]*University of British Columbia, School of Kinesiology, Vancouver, Canada*
[3]*Simon Fraser University, Burnaby, Canada*
[4]*University of British Columbia, Faculty of Education, Vancouver, Canada*
[5]*University of British Columbia, Department of Family Practice, Vancouver, Canada*

Correspondence to:

Johann Windt (JohannWindt@gmail.com)

Please cite as:

Windt, J., Hamilton, K., Zumbo, B. D., Cox, D N., & Sporer, B. (2021). Capturing the 'expert's eye': Towards a better understanding and implementation of subjective performance evaluations in team sports. https://doi.org/10.51224/SRXIV.6

Key Words: Performance evaluation, surveys, talent identification, validity theory

All authors have read and approved this version fo the manuscript. The article was last modified in August, 2021.

ABSTRACT

Subjective evaluations of athletic performance drive decision making across sporting organizations. Every day, based on their expertise and intuition, coaches select their starting lineups, scouts recommend or discourage teams from signing new potential players, and academy directors make decisions on which players move up or move out of a team's academy system. While this intuitive evaluation of performance occurs constantly, little attention has been given to how this process can be formally designed, implemented, and assessed to capture these expert evaluations of performance more effectively and thereby better inform decision making within sports organizations.

INTRODUCTION

Subjective evaluation of performance is widely accepted in sport. Informally, these assessments fuel conversations in which fans denounce 'the worst player they've ever seen' and they fill sports-radio airwaves over which pundits and talk-show guests declare their opinions on a team's merits or failings. More formally, coaches' subjective evaluations of their roster largely guide decisions regarding starting line-up selections and subsequent playing time decisions; a scout's subjective opinion of a player is vital in the decision to sign an athlete; and an academy director's subjective evaluation of youth players can guide decisions around player promotion or dismissal from an academy system [1,2]. The validity and reliability of a subjective evaluation may vary dramatically depending on the qualifications and expertise of the person who performs the assessment (e.g. fan/parent vs. coach/scout), in the formality and 'stakes' of the assessment (e.g. living room conversation vs. radio talk show vs. lineup selection vs. player recruitment), and in the characteristics of the evaluation tool itself. While varied and diverse, it is clear that subjective evaluation is a cornerstone of the sporting world. Despite this, there has been little written in the sport science literature that details how these evaluations can be strategically captured and evaluated. Based on this foundational gap, we have articulated four primary aims:

1) Define "subjective performance evaluations"
2) Describe the potential benefits of subjective performance evaluations
3) Describe existing and potential uses of subjective performance evaluations within team sporting organizations
4) Connect with the academic and applied community to further understand how subjective evaluations may be developed and implemented effectively.

WHAT ARE "SUBJECTIVE PERFORMANCE EVALUATIONS"?

Broadly, we define a ***subjective performance evaluation*** as any measure that captures human perception of a sport performance construct.

First, **'performance evaluation'**. Sport performance is complex and can be challenging to evaluate. In some instances, performance can be quantified objectively – who can throw the javelin

the farthest? However, even in these 'simple' cases (e.g. single-athlete events, closed environments), the question of which athlete is 'better' may be complicated by factors like the athlete's consistency e.g. does one throw in training make an athlete the best?; what is their ability to perform in competition contexts under pressure; and how they compare to reasonable contextual counterparts e.g. athletes of the same sex and age. These challenges are compounded when we consider more dynamic, open environment sports like team invasion sports like football (soccer). Ultimately, sport performance is a multifaceted 'construct' that no single measure can capture [3], with underlying component constructs from the physical, technical, psychological, and tactical domains [4]. A performance evaluation then, is any measurement taken with the aim of understanding these component constructs or overall athlete performance (Table 1).

Second, 'subjective'. We do not want to infer that this measurement category is inferior to other 'objective' performance measures. Rather, we differentiate the primary origin of these measurements. Subjective measures capture some aspect of human perception, while objective measures are generally perception-agnostic.

Table 1: Differentiating 'subjective' and 'objective' performance measures.

	Subjective Performance Measure	**Objective Performance Measure**
Definition	An evaluation of an athlete's or team's performance based on human perception.	An evaluation of an athlete's or team's performance based on any measurable metric other than human perception.
Advantages	• Affordable • May capture complex or hard-to-quantify metrics (e.g. tactical performance) • May capture the 'expert's eye' • May encourage staff and player buy-in	• Usually less influenced by human bias. • May be automated and enable more seamless data collection with less time required from staff/players.
Disadvantages	• May be largely influenced by human bias. • May be challenging to ensure inter- and intra-tester reliability without scale consistency, education. • Care must be taken to ensure scores reflect performance instead of other confounding factors i.e. validity. • Requires time from staff or players to complete the evaluations. • Requires critical thought and time to design and evaluate measures.	• May be limited to their context (e.g. match event data cannot capture training performances, GPS data may not be allowed in match play) • No technology is perfect, so measurement error will influence scores on these outcomes. • Can be very difficult to evaluate some more complex performance constructs like mental performance and tactical execution without advanced modelling. • May be expensive and thus more difficult to implement at lower-resourced/grassroots clubs.

'Subjective performance evaluation' is a broad definition that encompasses many performance constructs, most of which will also have associated objective measures. Consider a soccer player's 'endurance' as a construct. Knowing a player's endurance may help a team to decide whether or not the player is ready to play a full 90-minute soccer match. Many will suggest that one can measure endurance 'objectively' in a laboratory setting (e.g. VO_2 max) [5]. However, we could also ask a fitness coach to "subjectively" rate an athlete's endurance on a scale [6]. These may not align perfectly. In the limited studies that have compared such measures, agreement between subjective and objective indicators are not very high [6]. In this context it may be tempting to immediately consider the laboratory test as superior. However, it is important to consider sport performance as a multifaceted construct and appreciate where a VO_2 max test fits into its evaluation (Figure 1) [7]. A VO_2 max test may be the gold standard for assessing an athlete's maximal oxygen uptake and utilization, but it does not necessarily correspond to their ability to utilize their fitness within match play. The VO_2 max test is also influenced by the athlete's motivation to perform maximally in a laboratory environment, often on a bike. In this example, a fitness coaches' subjective evaluation may be more contextually grounded in how the athlete performs in competition, potentially providing a free, less invasive, and in some ways a more valid endurance performance measure. We contend that the more complex the construct, the more appropriate a subjective measure may become. In this example, asking a fitness coach to evaluate 'endurance' makes more sense than asking the coach to predict an athlete's maximal oxygen utilization.

Figure 1 – A simplified visual breakdown of where endurance fits into the realm of overall athlete performance.

WHAT ARE SOME OF THE BENEFITS TO SUBJECTIVE PERFORMANCE MEASURES?

We believe that subjective performance evaluations are important, particularly in the context of team sports. Below we highlight four reasons for this opinion – two that focus on the type of data that can be captured, and two that revolve around their implementation benefits.

First, subjective performance evaluations may capture information about complex constructs that cannot be easily quantified by many traditional objective measures (e.g., tactical decision making and consistency, attitude, motivation, work ethic). An athletes' soft skills, i.e. the way they perform mentally under pressure, and their tactical understanding and commitment to the team's principles, may be evaluated by a coach who sees the way they compete. Existing objective data streams such as event/tracking data may not be appropriate or able to evaluate these constructs.

Second, subjective performance evaluations may form the basis of the expert's eye [1,8]. It is common to hear and see coaches' intuitions about athletes they believe 'have it'. Sometimes these athletes are deemed by a coach, scout, or other individual to have an 'x factor' that differentiates them from their peers, or that they have potential that is not yet realized. Subjective performance evaluations may capture these intuitions.

Third, subjective performance evaluations may be very affordable. From spreadsheets to open-source platforms, users can easily record evaluation responses for no-to-low cost. While many companies that provide objective performance data (e.g. GPS, tracking data) come with a hefty price tag, subjective performance evaluations may capture some analogous information, but at a fraction of the cost.

Finally, implementation science recognizes the importance of involvement from those involved in the process [9,10]. Engaging directly with practitioners and capturing their perceptions about performance constructs that they deem valuable for performance may encourage buy-in and excitement for implementation. Standardizing subjective performance evaluations also provides a common ground and common language for practitioners to speak and a platform in which to discuss their varying opinions.

HOW CAN SUBJECTIVE PERFORMANCE EVALUATIONS BE USED?

Subjective evaluations can be used to measure any relevant performance related constructs and by any individual who possesses adequate contextual knowledge. For example, they may be used to capture and quantify scouts' opinions as they evaluate potential future athletes, to evaluate academy staff's feelings toward athletes' future potential, to assess a coaches' perceptions of athletes' performance qualities (globally or in a specific skillset), and to enhance an athletes' self-assessments of their performance. Any team sport organization that consistently and carefully implements such subjective performance evaluations may be rewarded with a wealth of athlete, team, and practitioner-related data to inform decision making (Table 2) [11].

Table 2: Objective and subjective evaluation examples within several performance domains.

Performance Domain	Objective Example	Subjective Example
Overall	Valuing Actions by Estimating Probabilities [12]	Player Rankings (e.g. MVP vote)
	Inside Football Player Ratings [13]	Australian Football League Player Ratings [13]
Physical	Player maximum velocity	Fitness-coach ranked 'speed'. [6]
	Yo-Yo Intermittent Recovery Test [14] 30:15 Intermittent Fitness Test [15]	Fitness-coach rating of 'endurance' [6]
Technical	Pass completion over expected pass % [16,17]	Scout's rating of passing ability. [18]
Tactical	Formation adherence % [19]	Head coach evaluation of team's adherence to game model formation.
Mental	Player performance under pressure [20]	Coach's evaluation of a players attentional control and ability to perform in high pressure game moments.

Interestingly, in the field of athlete recovery and health, systematic review evidence suggests that 'subjective' self-reported measurements of athlete wellbeing and recovery status may outperform 'objective' measures (e.g. physiological/ biochemical indicators) [21]. There may be other performance-related situations in which subjective evaluation outperforms objective markers [22]. However, where available, subjective measures may best be viewed as a helpful complement to, rather than in place of, objective metrics [23].

WHAT QUESTIONS REMAIN?

As an overarching framework, both subjective and objective evaluations rely on the same overarching steps to reach an inference. 1) Data generation, together with 2) instrumentation, combine to produce a 'score', which 3) allows people to make an inference. Each of these steps have inherent limitations, assumptions, and potential pitfalls (Figure 2). Returning to our simple example of a $V0_2$ max test and fitness coach's assessment of player endurance, the data generating process for the $V0_2$ max test is the athlete riding the bike to exhaustion. If they fail to put in the maximum effort, or something else influences their ability to perform the test (e.g. they have a quad strain), the data generating process is amiss. In the subjective example, the data generating process occurs as the physical preparation coach considers the relevant factors related to the player's endurance. If they lack the adequate knowledge and context, or they focus on athlete characteristics that are unrelated to their endurance (e.g. the athlete was rude earlier that day), the

data generation process is compromised. Even if the data generation process is perfect, instrumentation is crucial in accurately capturing the score. The bike test can be ruined by poor bike calibration, a mask that doesn't fit properly, sensors that malfunction or an inadequate warm-up protocol. A subjective evaluation can likewise be negatively impacted by a poorly designed evaluation tool, through inappropriate item selection, delivery medium, timing, or other issue. In either case, the score produced by each test only approximates the construct in question, and the final step is the inference someone makes with the score.

We contend that the sport science literature as a whole has focused a lot of attention on objective performance markers, both in selecting the proper 'data generating' test, and the instrumentation required to perform these tests reliably. However, the same cannot be said about subjective performance markers. Although many inferences are made about player and team quality on the basis of subjective intuitions, we know comparatively little about the data generating process for these evaluations [8], and even less about the instrumentation of capturing these evaluations effectively.

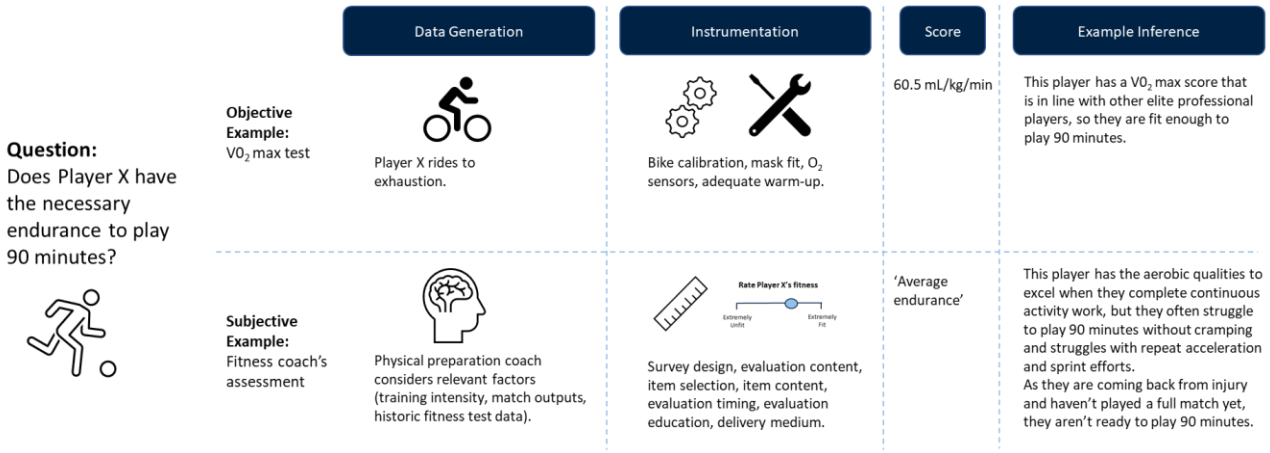

Figure 2: An overarching framework of how 'subjective' and 'objective' evaluations can be used to inform inferences.

Practically speaking, at the Vancouver Whitecaps Football club, we strive to make the most of subjective performance evaluations and capture the expertise of our practitioners, as a complement to our objective data streams. However, trying to implement these evaluations is fraught with questions. Consider a simple post-match subjective evaluation that a head and assistant coach commit to independently complete after each match. We could design such an evaluation in several ways – consider Figure 3 where we show 3 sample subjective items.

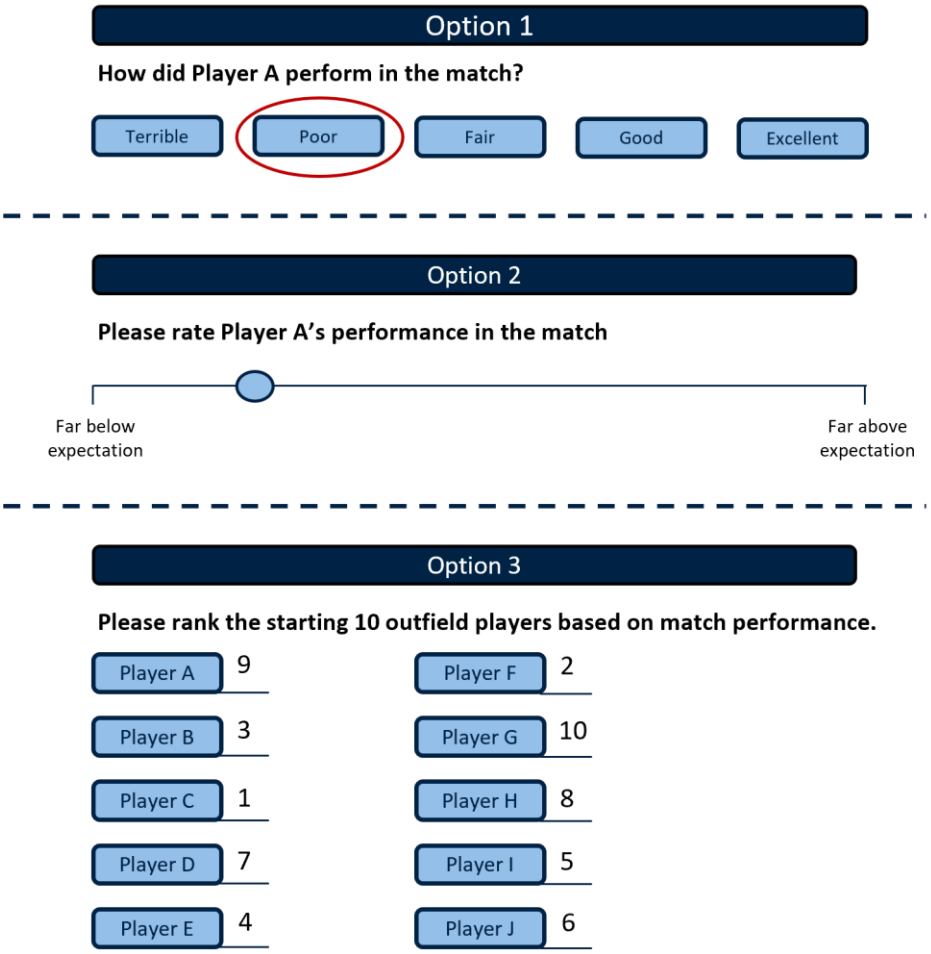

Figure 3 – Three example items that can all be used to capture coaches' subjective perceptions of each player's match performances. Although not shown, options 1 & 2 would need to completed for all players.

To our knowledge, few scales with validity and reliability-related evidence are widely available for coaches to evaluate player or team performances in this way. Where scales have been implemented, approaches have varied. Some have used player ranks for specific constructs like Option 3 [6,24], others have predicted future career level [25], and others have used categorical performance levels (e.g. bottom 25%, 25-50%, etc.) [24].

In the absence of well-established existing scales, or where existing scales do not focus on the constructs of interest, organizations may have to design their own. However, few resources exist in the sport science/performance literature on how a club can pragmatically develop their own subjective performance evaluations (i.e. instrumentation). For example, how should they answer questions like the following, which arise when creating items/evaluations like those in Figure 3:

- What content should we include in our evaluation, is this one item sufficient or are several necessary?
- Should items be categorical or continuous? If categorical or ordinal, how many categories are best?
- What are the anchors/poles (bottom/top points) that should be used on a given item?
- How should we deliver the evaluation? Electronically or digitally? What are the implications of each?
- How can we test an evaluation that we've developed to assess its key attributes: intra-rater reliability, inter-rater reliability, internal consistency, etc.
- How can we reduce the biases that commonly influence the accuracy and validity of human feedback [26,27]?

Answering each, and all, of these questions is beyond the scope of this paper, but we believe the measurement science and psychometric field can provide a pragmatic, guiding framework to develop and implement subjective evaluations in sports like it has elsewhere [28].

CALLS TO ACTION

Subjective performance evaluations are already informally implemented by virtually every sporting institution in the world, from grassroots to the professional level. Yet, few peer-reviewed subjective scales of team or individual sport performance exist for teams to adopt, and few resources are available to help organizations develop and instrument their own tools in a principled, pragmatic way. Therefore, to the academic community – we call for scales to be developed, validated, and disseminated to the applied sporting community. Moreover, acknowledging that academically produced scales may not meet the needs of each individual club, we call practitioners to consider how psychometric principles may guide sporting organizations looking to develop and implement effective subjective performance evaluations.

References

1 Jokuschies N, Gut V, Conzelmann A. Systematizing coaches' 'eye for talent': Player assessments based on expert coaches' subjective talent criteria in top-level youth soccer. *International Journal of Sports Science & Coaching* 2017;**12**:565–76. doi:10.1177/1747954117727646

2 Williams AM, Reilly T. Talent identification and development in soccer. *Journal of Sports Sciences* 2000;**18**:657–67. doi:10.1080/02640410050120041

3 Atkinson G. Sport performance: variable or construct? *Journal of Sports Sciences* 2002;**20**:291–2. doi:10.1080/026404102753576053

4 Larkin P, O'Connor D. Talent identification and recruitment in youth soccer: Recruiter's perceptions of the key attributes for player recruitment. *PLOS ONE* 2017;**12**:e0175716. doi:10.1371/journal.pone.0175716

5 Slimani M, Znazen H, Miarka B, *et al.* Maximum Oxygen Uptake of Male Soccer Players According to their Competitive Level, Playing Position and Age Group: Implication from a Network Meta-Analysis. *J Hum Kinet* 2019;**66**:233–45. doi:10.2478/hukin-2018-0060

6 Dugdale JH, Sanders D, Myers T, *et al.* A case study comparison of objective and subjective evaluation methods of physical qualities in youth soccer players. *Journal of Sports Sciences* 2020;**38**:1304–12. doi:10.1080/02640414.2020.1766177

7 Impellizzeri FM, Marcora SM. Test Validation in Sport Physiology: Lessons Learned from Clinimetrics. *International Journal of Sports Physiology and Performance* 2009;**4**:269–77. doi:10.1123/ijspp.4.2.269

8 Christensen MK. "An Eye for Talent": Talent Identification and the "Practical Sense" of Top-Level Soccer Coaches. *Sociology of Sport Journal* 2009;**26**:365–82. doi:10.1123/ssj.26.3.365

9 Durlak JA, DuPre EP. Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation. *Am J Community Psychol* 2008;**41**:327–50. doi:10.1007/s10464-008-9165-0

10 Saw AE, Main LC, Gastin PB. Monitoring Athletes Through Self-Report: Factors Influencing Implementation. *J Sports Sci Med* 2015;**14**:137–46.

11 Ward P, Windt J, Kempton T. Business Intelligence: How Sport Scientists Can Support Organisation Decision Making in Professional Sport. *International Journal of Sports Physiology and Performance* 2019;:1–10. doi:10.1123/ijspp.2018-0903

12 Decroos T, Bransen L, Van Haaren J, *et al.* Actions Speak Louder than Goals: Valuing Player Actions in Soccer. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Anchorage, AK, USA: : Association for Computing Machinery 2019. 1851–61. doi:10.1145/3292500.3330758

13 McIntosh S, Kovalchik S, Robertson S. Comparing subjective and objective evaluations of player performance in Australian Rules football. *PLOS ONE* 2019;**14**:e0220901. doi:10.1371/journal.pone.0220901

14 Bangsbo J, Iaia FM, Krustrup P. The Yo-Yo intermittent recovery test : a useful tool for evaluation of physical performance in intermittent sports. *Sports Med* 2008;**38**:37–51.

15 Buchheit M. The 30-15 intermittent fitness test: accuracy for individualizing interval training of young intermittent sport players. *J Strength Cond Res* 2008;**22**:365–74.

16 Spencer B, Jackson K, Bedin T, *et al.* Modelling the quality of player passing decisions in Australian Rules football relative to risk, reward and commitment. *Front Psychol* 2019;**10**. doi:10.3389/fpsyg.2019.01777

17 Fernández J, Barcelona FC, Bornn L, *et al.* Decomposing the Immeasurable Sport: A deep learning expected possession value framework for soccer. 2019;:20.

18 Bergkamp TLG, Frencken WGP, Niessen ASM, *et al.* How soccer scouts identify talented players. *European Journal of Sport Science* 2021;**0**:1–11. doi:10.1080/17461391.2021.1916081

19 Shaw L, Glickman M. Dynamic analysis of team strategy in professional football. ;:13.

20 Bransen L, Robberechts P, Van Haaren J, *et al.* Choke or Shine? Quantifying Soccer Players' Abilities to Perform Under Mental Pressure. *MIT Sloan Sports Analytics Conference* 2019.

21 Saw AE, Main LC, Gastin PB. Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review. *Br J Sports Med* 2016;**50**:281–91. doi:10.1136/bjsports-2015-094758

22 Crowcroft S, Slattery K, McCleave E, *et al.* Do Athlete Monitoring Tools Improve a Coach's Understanding of Performance Change? *International Journal of Sports Physiology and Performance* 2020;**15**:847–52. doi:10.1123/ijspp.2019-0338

23 Sieghartsleitner R, Zuber C, Zibung M, *et al.* Science or Coaches' Eye? – Both! Beneficial Collaboration of Multidimensional Measurements and Coach Assessments for Efficient Talent Selection in Elite Youth Football. *J Sports Sci Med* 2019;**18**:32–43.

24 McCormack S, Jones B, Elliott D, *et al.* Coaches' Assessment of Players Physical Performance: Subjective and Objective Measures are needed when Profiling Players. *European Journal of Sport Science* 2021;:1–11. doi:10.1080/17461391.2021.1956600

25 Cripps AJ, Hopper LS, Joyce C. Can coaches predict long-term career attainment outcomes in adolescent athletes? *International Journal of Sports Science & Coaching* 2019;**14**:324–8. doi:10.1177/1747954119848418

26 Montibeller G, von Winterfeldt D. Cognitive and Motivational Biases in Decision and Risk Analysis: Biases in Decision and Risk Analysis. *Risk Analysis* 2015;**35**:1230–51. doi:10.1111/risa.12360

27 Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974;**185**:1124–31. doi:10.1126/science.185.4157.1124

28 Burton LJ, Mazerolle SM. Survey instrument validity part I: Principles of survey instrument development and validation in athletic training education research. *Athletic Training Education Journal* 2011;**6**:27–35.