This is an Accepted Manuscript of an article published by Taylor & Francis Group in *Journal of Sports Sciences* on 13 Jun 2020, available online: http://www.tandfonline.com/10.1080/02640414.2020.1766177

Title of the article:

A case study comparison of objective and subjective evaluation methods of physical qualities in youth soccer players

Submission type:

Original article

Author names:

James H. Dugdale[1], Dajo Sanders[2], Tony Myers[3], A. Mark Williams[4], Angus M. Hunter[1]

Institutional affiliations:

[1] Physiology, Exercise and Nutrition Research Group, Faculty of Health Sciences and Sport, University of Stirling, Scotland, United Kingdom – a.m.hunter1@stir.ac.uk

[2] Department of Human Movement Science, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands – dajosanders@gmail.com

[3] Sport, Physical Activity and Health Research Centre, Newman University, Birmingham, United Kingdom – tony.myers@newman.ac.uk

[4] Department of Health, Kinesiology, and Recreation, University of Utah, Salt Lake City, UT, USA – mark.williams@health.utah.edu

Corresponding author:

James H. Dugdale
Faculty of Health Sciences and Sport - University of Stirling, Stirling, Scotland, FK9 4LA.
t. +44 (0)1786 466290 e. james.dugdale@stir.ac.uk

Preferred Running Head:

Objective and subjective ratings in youth soccer

Abstract Word Count:

222

Text-only Word Count:

4,986

Number of Figures:	**Number of Tables:**
2	2

A CASE STUDY COMPARISON OF OBJECTIVE AND SUBJECTIVE EVALUATION METHODS OF PHYSICAL QUALITIES IN YOUTH SOCCER PLAYERS

Abstract

Subjective and objective assessments may be used congruently when making decisions regarding player recruitment in soccer, yet there have been few attempts to examine the level of agreement between these methods. Therefore, we compare levels of agreement between subjective and objective assessments of physical qualities associated with youth soccer performance. In total, 80 male youth soccer players (13.2 ± 1.9 years), and 12 professional coaches volunteered to participate. Players were objectively assessed using five fitness measures: Yo-Yo Intermittent Recovery Test Level 1; Countermovement vertical jump; Functional Movement Screen™; 5/20m sprint; alongside anthropometric measures. Additionally, coaches subjectively rated each player on the same five physical qualities using 5-point Likert scales. Inter-rater agreement between ratings from lead and assistant coaches were established for each age group. Moreover, Bayesian regression models were fitted to determine how well coach ratings were able to predict fitness test performance. Although inter-rater agreement between lead and assistant coaches was moderate-to-substantial (ω=0.48-0.68), relationships between coaches subjective rating's and corresponding fitness test performance were only highly related for the highest and lowest performing players. We suggest that while ratings derived from objective and subjective assessment methods may be related when attempting to differentiate between distinct populations, concerns exist when evaluating homogeneous samples using these methods. Our data highlight the benefits of using both types of measures in the talent identification process.

Key words: Coach ratings; fitness testing; talent identification; perception; adolescent.

Introduction

Identifying and developing talented young athletes is integral to the coach's role in soccer (Larkin & O'Connor, 2017; Reeves, Roberts, McRobert, & Littlewood, 2018; Reilly, Williams, Nevill, & Franks, 2000; Williams & Reilly, 2000). Traditionally, clubs have employed scouting systems where coaches view players in a training or game scenario and assess them based on their perceived performance and ability (Unnithan, White, Georgiou, Iga, & Drust, 2012; Williams & Reilly, 2000). However, if used in isolation, these processes may lead to potentially biased results (Meylan, Cronin, Oliver, & Hughes, 2010). During their development, youth soccer players may encounter several coaches, each with varying conscious or unconscious philosophical and cognitive biases (Unnithan *et al*., 2012). Nonetheless, experiential knowledge gathered from coaching, playing, and scouting continues to carry substantial weight in decision making when prescribing training programmes and when players are selected into (or deselected from) systematic training structures (Grossmann & Lames, 2015; Musculus & Lobinger, 2018).

Scientists have attempted to better understand the potential attributes and strategies used by coaches and recruiters during talent identification and development (Hendry, Williams, & Hodges, 2018; Larkin & O'Connor, 2017; Reeves, McRobert, Lewis, & Roberts, 2019; Reeves, Roberts, *et al*., 2018). From an Australian perspective, Larkin and O'Connor (2017) reported a range of technical, tactical, physiological, and psychological parameters perceived by experienced professional youth soccer coaches to be "key attributes" for entry level recruitment. Similarly, Roberts, McRobert, Lewis, and Reeves (2019) presented a UK perspective, exploring both generic and position-specific attributes that may be important to progression in youth soccer. The results from these studies encourage the use of multi-disciplinary and player-positional attributes during the talent identification process, while acknowledging that physiological and anthropometric qualities may be less important to coaches when selecting junior-elite youth players. In contrast to these studies, there is a plethora of work spanning the last 20 years suggesting that objectively assessed physical abilities may be an important contributing factor related to recruitment, selection, and progression from youth to senior level in soccer.

For example, elite soccer players are greater in physical stature and mass, and perform better on sprint, endurance, strength, and power assessments compared to players of a lower playing standard (Dugdale, Arthur, Sanders, & Hunter, 2019; Gil, Ruiz, Irazusta, Gil, & Irazusta, 2007; Rebelo *et al*., 2013). Similarly, physical qualities have been suggested to discriminate between players retained or released within a soccer academy, and when evaluating successful vs. unsuccessful academy graduation (Emmonds, Till, Jones, Mellis, & Pears, 2016; Figueiredo, Gonçalves, Coelho e Silva, & Malina, 2009; le Gall, Carling, Williams, & Reilly, 2010). Consequently, physical and physiological testing have become common methods within applied practice and field-based research in an effort to provide a more substantive reference base of key physical qualities underpinning player development (Enright *et al*., 2018; Pyne, Spencer, & Mujika, 2014), and talent identification in soccer (Dugdale *et al*., 2019; Murr, Raabe, & Höner, 2018). However, because of the complex and multifaceted nature of soccer, these data may be limited in their prognostic ability (Bergkamp, Niessen, Den Hartigh, Frencken, & Meijer, 2019; Murr, Raabe, *et al*., 2018; Roberts *et al*., 2019). The need to adopt a more holistic approach to talent identification and development, accompanying objective measures with traditional subjective decision making processes, has been widely endorsed in youth soccer (Bergkamp *et al*., 2019; Höner & Votteler, 2016; Murr, Feichtinger, Larkin, O’Connor, & Höner, 2018; Sieghartsleitner, Zuber, Zibung, & Conzelmann, 2019; Unnithan *et al*., 2012).

Only a select number of researchers have examined both objective and subjective measures congruently in soccer. Sieghartsleitner *et al*. (2019) examined both objective and subjective assessment methods from multiple dimensions across a prognostic period of five years (U14-U19) in an elite sample of players in Switzerland. Similarly, in their sample of highly trained pre-adolescent youth soccer players, Fenner, Iga, and Unnithan (2016) examined small-sided game assessments as a viable talent identification tool through the unification of objective and subjective measurements. The results from these studies suggest that while subjective coach assessments are likely to be holistic in nature involving the integration of multiple game-based aspects simultaneously, the addition of objective data to support subjective coach assessment methods may improve prognostic ability during talent identification.

Despite the increasing interest in complementing subjective assessments with objective data, when examining physical predictors within talent identification and development in soccer, the majority of researchers have only estimated relationships between physical qualities and performance criteria (Deprez, Fransen, Lenoir, Philippaerts, & Vaeyens, 2015; Gonaus & Müller, 2012; Höner & Feichtinger, 2016; Höner & Votteler, 2016). As a consequence, more empirical work is needed to better identify how subjective and objective assessments of physical qualities in soccer players are related, and, the extent to which the use of subjective judgements of physical qualities, in their own right, may be justified.

In the current study, we had two aims. First, we examined the relationship between subjective coach ratings for a range of physical qualities previously reported as relevant to successful performance in soccer, with a corresponding objective measure of the same component of physical fitness. Second, we examined the inter-rater agreement between two coaches (lead vs. assistant) when subjectively rating youth players on a range of physical abilities relative to successful performance in soccer.

Methods

Participants

Players

In total, 80 male youth soccer players aged 10.2 to 16.7 years (M: 13.2 ± 1.9) were recruited. Player stature ranged from 130.1 to 185.3 cm (M: 160.3 ± 13.9), and player mass ranged from 27.4 to 83.7 kg (M: 49.3 ± 12.4). We used an exploratory case study design (Reeves *et al*., 2019; Yin, 2009) using players affiliated to a junior-elite soccer academy playing at the highest competitive level in Scotland. Participants were categorised into age groups as specified by the Scottish Football Association (SFA): U11 (*n*=16); U12 (*n*=14); U13 (*n*=11); U14 (*n*=12); U15 (*n*=12); and U17 (*n*=15). We obtained informed assent from all participants, consent from parents/guardians, and gatekeeper consent from the Academy Director prior to collecting data. The study received institutional ethical approval (GUEP 533R).

Coaches

We recruited twelve male soccer coaches. The lead and assistant coach for each of the six age groups listed above were recruited for the study. The Lead Coaches ranged from 29.6 to 55.8 years (M: 40.5 ± 10.2) of age, and their coaching experience ranged from 6.25 to 20.0 years (M: 13.5 ± 5.7) with 0.5 to 4.0 years (M: 1.8 ± 1.4) coaching history with their current team. Lead Coaches held either the SFA Advanced Children's or the UEFA Youth A licence coaching qualifications. The Assistant Coaches ranged from 23.3 to 55.0 years (M: 37.8 ± 13.7) of age, and their coaching experience ranged from 4.0 to 20.0 years (M: 13.3 ± 6.5) with 0.5 to 2.0 years (M: 1.3 ± 0.8) coaching history with their current team. The coaching qualifications held by Assistant Coaches ranged from no formal coaching qualification to the UEFA Youth B licence coaching qualification. We obtained informed consent from all coaches prior to data collection.

Procedures

Fitness Tests

We collected objective data on five measures of physical fitness using established methods: Yo-Yo Intermittent Recovery Test Level 1 (YYIRT L1) (Krustrup *et al*., 2003); countermovement vertical jump (CMJ) (Murtagh *et al*., 2018); Functional Movement Screen™ (FMS) (Cook, Burton, & Hoogenboom, 2006); and 5m/20m linear sprint tests (Enright *et al*., 2018). Moreover, we recorded body mass, stature, and seated height. A regression equation was used to provide somatic maturity estimates, presented as maturity offset (years from age at peak height velocity) (Mirwald, Baxter-Jones, Bailey, & Beunen, 2002). The fitness tests selected are commonly used as generic physical fitness tests within a youth soccer population (Paul & Nassis, 2015), as well as being appropriate for implementation across the entire age range of the selected sample (Dugdale *et al*., 2019; Gil, Gil, Ruiz, Irazusta, & Irazusta, 2007). Also, the physical qualities measured have been reported to be desirable in elite adult players (Dodd & Newans, 2018).

The testing sessions were completed a minimum of 48 hours following a competitive game, and in absence of strenuous exercise within 24 hours prior. The testing sessions were conducted indoors (~22°C) on a non-slip sports hall playing surface. Participants conducted a standardised warm-up protocol consisting of light aerobic activity, dynamic stretching, and progressive sprinting. Following the standardised warm-up, participants received verbal instructions and demonstrations from the research team immediately prior to conducting three familiarisation attempts for each test. Guidance and feedback were provided to participants by the research team following each familiarisation attempt, however no guidance was provided to participants between recorded attempts. Participants completed three attempts of each test (with exception of the YYIRT L1) with the best attempt being selected for analysis. We standardised recovery intervals at three minutes for each test.

Coach ratings

We collected subjective data on the qualities intended for assessment by the physical fitness tests. The physical qualities rated by the coaches were: 'Endurance' – YYIRT L1; 'Power' – CMJ; 'Movement Quality' – FMS™; 'Physical Development' – maturity offset; 'Acceleration' – 5m linear sprint; and 'Sprint Speed' – 20m linear sprint. Coaches used a 5-point Likert scale to rate the physical abilities of each player: 1 – poor; 2 – below average; 3 – average; 4 – very good; and 5 – excellent. Such coach-based rating methods have previously been adopted by researchers and they demonstrate good reliability and validity (Ali, 2011; Hendry *et al*., 2018; Larkin & O'Connor, 2017; Unnithan *et al*., 2012). The Lead and Assistant Coach for each age group provided separate ratings for players from their squad at identical time points and using an identical scale. The coaches completed their subjective ratings before a regular scheduled training session, one week prior to players completing the fitness testing battery. Coach's ratings were completed independently without confirmation with other coaches or support staff.

Statistical Analysis

We present descriptive statistics of physical test performance associated with Lead and Assistant Coach ratings of corresponding subjective qualities as means and standard deviations (SD). Inter-rater agreement between the Lead and Assistant Coach is reported as Sklar's ω and interpreted as: ($\omega \leq 0.2$) – slight agreement; ($0.21 < \omega \leq 0.4$) – fair agreement;

(0.41 < $\omega \leq 0.6$) – moderate agreement; (0.61 < $\omega \leq 0.8$) – substantial agreement; ($\omega > 0.81$) – near-perfect agreement (Hughes, 2018). A series of Bayesian regression models were fitted to determine how well coach ratings predict performance in measures assessing corresponding physical qualities. Leave-One-Out cross-validation (LOO) was used to determine the best model for predicting relationships between ratings and measured variables. LOO is a method of estimating pointwise out-of-sample prediction accuracy from fitted Bayesian models using log-likelihoods from posterior simulations of the parameter values (Vehtari, Gelman, & Gabry, 2017). The best models, those with the lowest LOO information criterion, were Bayesian monotonic ordinal regression models.

Bayesian monotonic ordinal regression models allow ordinal predictors to be modelled without falsely treating them either as continuous or as unordered categorical predictors, meaning predictors may be non-equidistant with respect to their relationship to a response variable. For example, coach ratings on a 5-point scale (1 = poor to 5 = excellent) cannot be considered interval level data. While they have a meaningful order, the intervals between ratings may be uneven. Therefore, while a rating of four is higher than a rating of one, two or three, it is not twice the value of two. Treating ordinal ratings as if they were on an interval scale can lead to inaccurate predictions and inaccurate relationships. We present estimates from the models along with 95% credible intervals and associated simplex parameters. We analysed the data via R (R Core Team, 2018) using the sklarsomega package to calculate Sklar's ω and the brms package (Bürkner, 2018) to fit all the Bayesian models. Brms uses Stan (Stan Development Team, 2018) to implement a Hamiltonian Markov Chain Monte Carlo (MCMC) with a No-U-Turn Sampler. All models were checked for convergence ($\hat{r}$ = 1), with the graphical posterior predictive checks showing simulated data under the best fitted models compared well to the observed data with no systematic discrepancies (Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2019).

Results

Predictive ability of coach subjective ratings relative to fitness test performance

The descriptive data from measured variables for the ratings provided by each coach and the corresponding physical abilities are presented in **Table 1**. The Bayesian monotonic ordinal regression models show the ratings awarded by both the Lead and Assistant Coaches are not evenly assigned when compared to objectively measured performance (**Figure 1**). Visual inspection shows the data are skewed for different rating categories across measures. The marginal effects for the Bayesian monotonic ordinal regression models show that the ratings by both the Lead and Assistant Coach have nonlinear relationships with the measured variables predicted (**Figure 2**).

(Table 1 about here)

(Figure 1 about here)

(Figure 2 about here)

Inter-rater reliability and accuracy of coach subjective ratings

The Lead and Assistant Coach ratings displayed moderate ($0.41 < \omega \leq 0.6$) to substantial ($0.61 < \omega \leq 0.8$) agreement when rating physical abilities on a 5-point rating scale (**Table 2**). The ratings provided by the Lead Coach explained a higher percentage of variance in performance variables across models than those awarded by the Assistant Coach (**Table 2**). Variance explained differed depending on the quality rated. The highest variance explained was the Lead Coach's ratings for endurance which explained 23% of the variance in the YYIRT L1. The lowest variance explained was 1% of the variance in FMS™, explained by the Assistant Coach's ratings of movement quality (**Table 2**). The Lead Coach's highest ratings equated to the best performances for YYIRT L1, CMJ, FMS, 5m and 20m sprint. The lowest ratings awarded by the Lead Coach equated to the poorest performances for CMJ, 5m and 20m sprint. However, the only variable where the Lead Coaches progressively higher ratings align with a progressively better mean performance was for CMJ performance (**Table 1**). The Assistant Coaches highest ratings equated to the best performances for CMJ, 5m and 20m sprint, and the lowest ratings to the poorest performances for YYIRT L1, FMS and 5m sprint. The only variable where mean performances increase with progressively higher ratings by the Assistant Coach is for 5m sprint performance (**Table 1**).

(Table 2 about here)

Discussion

Our results indicate that levels of agreement between objective (fitness test performance) and subjective (coach ratings) data on physical qualities were skewed in nature and displayed different levels of variance across tests. Although coaches exhibited accuracy when providing ratings for lowest/highest performers, explained variance between ratings scores (1-5) fluctuated, with no consistent trend observed across physical qualities for Lead and Assistant Coaches. Also, while Lead and Assistant Coaches displayed moderate-to-substantial agreement in their ratings of perceived physical qualities of players, the levels of agreement between them were the lowest (moderate) for 'endurance', and the highest (substantial) for 'power'.

Although coaches were particularly accurate when rating the highest and lowest performers, a substantial amount of variance in fitness test performance was observed between players allocated a moderate rating (2-4). The skewed nature of the data observed between coach rating and fitness test performance potentially supports the method of using coach-based rating/ranking procedures for talent identification processes, as coaches seem to be able to correctly identify individuals at the extremities of a scale (lowest/highest) (Fenner *et al*., 2016; Reilly *et al*., 2000; Unnithan *et al*., 2012). However, our results highlight the subjective and potentially biased nature of coach rating systems, as well as their limitations, when trying to differentiate between performers of similar abilities (Meylan *et al*., 2010). Therefore, similar to emerging suggestions from relative age effect and maturation-selection phenomenon research (Reeves, Enright, Dowling, & Roberts, 2018), we encourage coaches and recruitment staff to be aware of this inability to differentiate between players at the extremities of these rating scales, and acknowledge the potential oversight that may be exhibited to those achieving “moderate” scores on objective and subjective measures.

Due to the complex and multi-faceted nature of soccer, researchers have suggested that reductionist and decontextualised testing may be inappropriate and that assessment of game-based activities may be more suitable (Bennett *et al*., 2018; Bergkamp *et al*., 2019; Unnithan *et al*., 2012). An argument could potentially be made to support this suggestion, considering we observed no consistent trend across ratings for physical qualities provided by Lead and Assistant Coaches. This questions the suitability of physical fitness tests to assess the key characteristics associated with successful performance in soccer. In our study, we acknowledge that disconnect may exist between the coaches perceptions of physical qualities (retrospective from in-situ performance) and objective assessments in an isolated and decontextualised setting. Therefore, we reiterate the importance of implementing contextual and game-based assessments within the talent identification process. Nonetheless, physical training and monitoring continues to be prioritised during the training process in soccer (Enright *et al*., 2018; Morgans, Orme, Anderson, & Drust, 2014). Considering the influence of coach subjective opinion during programme design and selection/deselection in soccer, our results suggest that coaches should consult objective data when making decisions regarding isolated physical qualities.

The moderate agreement observed between Lead and Assistant Coach ratings for “endurance” suggests that coaches may possess somewhat different perceptions of what constitutes poor-excellent endurance capacities. This discrepancy may be due to the intermittent nature of soccer and/or the multitude of exercise modalities and energy systems utilised within competition (Buchheit, Mendez-Villanueva, Simpson, & Bourdon, 2010;

Saward, Morris, Nevill, Nevill, & Sunderland, 2016). It has been suggested that "endurance" comprises of various facets including both aerobic and anaerobic capacities (Bangsbo, Mohr, & Krustrup, 2006; Stølen, Chamari, Castagna, & Wisloff, 2005). Consequently, multiple different procedures are implemented to assess the repeated and intermittent nature of performance in soccer (Buchheit, 2008; Hill-Haas, Dawson, Impellizzeri, & Coutts, 2011; Krustrup *et al*., 2003, 2006). This ambiguity regarding endurance capacity could therefore distract from a cohesive inter-rater perception and rating of this ability. We propose that the term "endurance" may be too vague, and that in future, a range of different physical qualities could be assessed capturing the multiple exercise modalities and energy systems exhibited during soccer.

In contrast, perceptions of "power", "acceleration", and "speed" displayed substantial agreement between coaches, suggesting that these qualities are more universally identifiable during soccer game-based activity. Soccer players playing at a higher competitive level often outperform those playing at a lower competitive level on tests related to "power" (eg. Dugdale *et al*., 2019), "acceleration" (eg. Coelho E Silva *et al*., 2010), and "speed" (eg. le Gall *et al*., 2010). Furthermore, specific positions may favour such physical qualities resulting in more obvious demonstrations of these qualities during performance for these players (Roberts *et al*., 2019). Our sample were recruited from a junior-elite academy and were likely highly trained along with holding a greater understanding of position-specific criteria for their stage of development (Roberts *et al*., 2019). An awareness of the relationships between these physical qualities and playing standard/position by coaches could, therefore, make them easier to identify during game-based activity (Reeves, Enright, *et al*., 2018; Roberts *et al*., 2019). Lastly, these physical qualities largely rely on neuromuscular factors (Stølen *et al*., 2005) and, as a result, are most affected by growth and maturation (Philippaerts *et al*., 2006).

Those with an advanced maturity status may demonstrate vastly different abilities on these qualities compared to late developers, which may be identified by coaches (Carling, Le Gall, & Malina, 2012; Reeves, Enright, *et al*., 2018). Our results suggest that these physical qualities may be easily detectable during game-based activity, and we encourage coaches to be aware of the potential influence that playing standard, playing position, and maturity status may have on the accuracy of their ratings.

Finally, we must acknowledge that the Lead Coaches within our sample were older, having gained more general coaching experience and accumulated more time coaching with the players that they rated during our study. General and group-specific experience gathered during a coach's career is suggested to influence quality of decision making and judgements in youth soccer (Cushion, Ford, & Williams, 2012). However, in our sample, these differences, when compared to the Assistant Coaches, were small. Nevertheless, we cannot rule out the possibility that this additional coaching exposure may have improved the accuracy of coach ratings for the Lead Coaches in our sample. We also observed that Lead Coaches held a higher level of formal coaching qualification compared to Assistant Coaches, some of whom held no formal coaching qualifications at all. While formal qualifications are rarely identified when assessing attributes of importance for soccer coaching (Reeves, Roberts, *et al*., 2018), they are often a prerequisite when coaching in an academy setting when working with junior-elite players. Given our study design, a more comprehensive knowledge of supplementary attributes related to performance (such as physical qualities) may have been experienced during more formal and structured learning, leading to more informed ratings by lead coaches. In future, we encourage researchers to consider the impact that coach experience and qualifications may have when collecting coach subjective ratings.

Our results should be considered in light of a number of limitations. First, this was exploratory adopting a single club case study design. We suggest that results are treated with appropriate caution given the design utilised. It has been established that clubs may adopt a specific philosophy, favouring various styles of play (Cobb, Unnithan, & McRobert, 2018; Williams & Reilly, 2000). Moreover, there is a tendency for coaches and practitioners to favour physical and anthropometric characteristics rather than technical capacities of young players (Reeves, Enright, *et al*., 2018; Reeves, Roberts, *et al*., 2018; Unnithan *et al*., 2012). Consequently, certain physical qualities, within our study, may have been rated by coaches under the influence of conscious or unconscious bias. The physical qualities assessed within our study develop at different times and rates throughout adolescence (Malina *et al*., 2005) and may be perceived to vary in importance across different playing positions (Roberts *et al*., 2019). Therefore, specific playing position, age group or maturity status analysis may provide a more comprehensive understanding of subjective ratings for these sub-groups. In future, the use of larger samples, spanning multiple clubs, may help negate concerns and extend our understanding of the complex relationships between subjective, coach-based ratings and objective, empirical tests.

In summary, the translation between objective and subjective assessment methods of physical qualities in youth soccer players may be effective when attempting to differentiate between distinct population groups. However, when evaluating homogeneous samples, these methods may lack sensitivity. A strong case exists to use both subjective and objective assessments in an integrated manner when attempting to identify strengths and weaknesses in youth soccer players.

Disclosure statement

The authors report no conflict of interest.

References

Ali, A. (2011). Measuring soccer skill performance: A review. *Scandinavian Journal of Medicine and Science in Sports*, *21*(2), 170–183. https://doi.org/10.1111/j.1600-0838.2010.01256.x

Bangsbo, J., Mohr, M., & Krustrup, P. (2006). Physical and metabolic demands of training and match-play in the elite football player. *Journal of Sports Sciences*, *24*(7), 665–674. https://doi.org/10.1080/02640410500482529

Bennett, K. J. M., Novak, A. R., Pluss, M. A., Stevens, C. J., Coutts, A. J., & Fransen, J. (2018). The use of small-sided games to assess skill proficiency in youth soccer players: a talent identification tool. *Science and Medicine in Football*, *2*(3), 231–236. https://doi.org/10.1080/24733938.2017.1413246

Bergkamp, T. L. G., Niessen, A. S. M., Den Hartigh, R. J. R., Frencken, W. G. P., & Meijer, R. R. (2019). Methodological Issues in Soccer Talent Identification Research: A Critical Review. *Sports Medicine*, *8*(6). https://doi.org/10.1007/s40279-019-01113-w

Buchheit, M. (2008). The 30-15 Intermittent Fitness Test: Accuracy for Individualizing Interval Training of Young Intermittent Sport Players. *Journal of Strength and Conditioning Research*, *22*(2), 365–374. https://doi.org/10.3141/1747-08

Buchheit, M., Mendez-Villanueva, A., Simpson, B. M., & Bourdon, P. C. (2010). Match Running Performance and Fitness in Youth Soccer. *International Journal of Sports Medicine*, *31*(11), 818–825. https://doi.org/10.1055/s-0030-1262838

Bürkner, P. C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/https://doi:10.32614/RJ-2018-017

Carling, C., Le Gall, F., & Malina, R. M. (2012). Body size, skeletal maturity, and functional characteristics of elite academy soccer players on entry between 1992 and 2003. *Journal of Sports Sciences*, *30*(15), 1683–1693. https://doi.org/10.1080/02640414.2011.637950

Cobb, N. M., Unnithan, V., & McRobert, A. P. (2018). The validity, objectivity, and reliability of a soccer-specific behaviour measurement tool. *Science and Medicine in Football*, *2*(3), 1–7. https://doi.org/10.1080/24733938.2017.1423176

Coelho E Silva, M. J., Figueiredo, A. J., Simïes, F., Seabra, A., Natal, A., Vaeyens, R., … Malina, R. M. (2010). Discrimination of U-14 soccer players by level and position. *International Journal of Sports Medicine*, *31*(11), 790–796. https://doi.org/10.1055/s-0030-1263139

Cook, G., Burton, L., & Hoogenboom, B. (2006). Pre-participation screening: The use of fundamental movements as an assessment of function - part 1. *North American Journal of Sports Physical Therapy*, *1*(2), 62–72.

Cushion, C., Ford, P. R., & Williams, A. M. (2012). Coach behaviours and practice structures in youth soccer: Implications for talent development. *Journal of Sports Sciences*, *30*(15), 1631–1641. https://doi.org/10.1080/02640414.2012.721930

Deprez, D. N., Fransen, J., Lenoir, M., Philippaerts, R. M., & Vaeyens, R. (2015). A retrospective study on anthropometrical, physical fitness, and motor coordination characteristics that influence dropout, contract status, and first-team playing time in high-level soccer players aged eight to eighteen years. *Journal of Strength and Conditioning Research*, *29*(6), 1692–1704. https://doi.org/10.1519/JSC.0000000000000806

Dodd, K. D., & Newans, T. J. (2018). Talent identification for soccer: Physiological aspects. *Journal of Science and Medicine in Sport*, *21*(10), 1073–1078. https://doi.org/10.1016/j.jsams.2018.01.009

Dugdale, J. H., Arthur, C. A., Sanders, D., & Hunter, A. M. (2019). Reliability and validity of field-based fitness tests in youth soccer players. *European Journal of Sport Science*, *19*(6), 745–756. https://doi.org/10.1080/17461391.2018.1556739

Emmonds, S., Till, K., Jones, B., Mellis, M., & Pears, M. (2016). Anthropometric, speed and endurance characteristics of English academy soccer players: Do they influence obtaining a professional contract at 18 years of age? *International Journal of Sports Science and Coaching*, *11*(2), 212–218. https://doi.org/10.1177/1747954116637154

Enright, K., Morton, J., Iga, J., Lothian, D., Roberts, S., & Drust, B. (2018). Reliability of "in-season" fitness assessments in youth elite soccer players: a working model for practitioners and coaches. *Science and Medicine in Football*, *2*(3), 177–183. https://doi.org/10.1080/24733938.2017.1411603

Fenner, J. S. J., Iga, J., & Unnithan, V. (2016). The evaluation of small-sided games as a talent identification tool in highly trained prepubertal soccer players. *Journal of Sports Sciences*, *34*(20), 1983–1990. https://doi.org/10.1080/02640414.2016.1149602

Figueiredo, A. J., Gonçalves, C. E., Coelho e Silva, M. J., & Malina, R. M. (2009). Characteristics of youth soccer players who drop out, persist or move up. *Journal of Sports Sciences*, *27*(9), 883–891. https://doi.org/10.1080/02640410902946469

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistics Society*, *182*(2), 389–402.

https://doi.org/10.1111/rssa.12378

Gil, S. M., Gil, J., Ruiz, F., Irazusta, A., & Irazusta, J. (2007). Physiological and anthropometric characteristics of young soccer players according to their playing position: Relevance for the selection process. *Journal of Strength and Conditioning Research*, *21*(2), 438–445. https://doi.org/10.1519/R-19995.1

Gil, S. M., Ruiz, F., Irazusta, A., Gil, J., & Irazusta, J. (2007). Selection of young soccer players in terms of anthropometric and physiological factors. *Journal of Sports Medicine and Physical Fitness*, *47*, 25–32.

Gonaus, C., & Müller, E. (2012). Using physiological data to predict future career progression in 14- to 17-year-old Austrian soccer academy players. *Journal of Sports Sciences*, *30*(15), 1673–1682. https://doi.org/10.1080/02640414.2012.713980

Grossmann, B., & Lames, M. (2015). From Talent to Professional Football – Youthism in German Football. *International Journal of Sports Science & Coaching*, *10*(6), 1103–1113. https://doi.org/10.1260/1747-9541.10.6.1103

Hendry, D. T., Williams, A. M., & Hodges, N. J. (2018). Coach ratings of skills and their relations to practice, play and successful transitions from youth-elite to adult-professional status in soccer. *Journal of Sports Sciences*, *36*(17), 2009–2017. https://doi.org/10.1080/02640414.2018.1432236

Hill-Haas, S. V., Dawson, B., Impellizzeri, F. M., & Coutts, A. J. (2011). Physiology of Small-Sided Games Training. *Journal of Sports Medicine*, *41*(3), 199–220. https://doi.org/10.2165/11539740-000000000-00000

Höner, O., & Feichtinger, P. (2016). Psychological talent predictors in early adolescence and their empirical relationship with current and future performance in soccer. *Psychology of Sport and Exercise*, *25*, 17–26. https://doi.org/10.1016/j.psychsport.2016.03.004

Höner, O., & Votteler, A. (2016). Prognostic relevance of motor talent predictors in early adolescence: A group- and individual-based evaluation considering different levels of achievement in youth football. *Journal of Sports Sciences*, *34*(24), 2269–2278. https://doi.org/10.1080/02640414.2016.1177658

Hughes, J. (2018). Sklar's Omega: A Gaussian copula-based framework for assessing agreement. *ArXiv E-Prints*.

Krustrup, P., Mohr, M., Amstrup, T., Rysgaard, T., Johansen, J., Steensberg, A., … Bangsbo, J. (2003). The Yo-Yo intermittent recovery test: Physiological response, reliability, and validity. *Medicine and Science in Sports and Exercise*, *35*(4), 697–705. https://doi.org/10.1249/01.MSS.0000058441.94520.32

Krustrup, P., Mohr, M., Nybo, L., Jensen, J. M., Nielsen, J. J., & Bangsbo, J. (2006). The Yo-Yo IR2 Test: Physiological Response, Reliability, and Application to Elite Soccer. *Medicine & Science in Sports & Exercise*, *38*(9), 1666–1673. https://doi.org/10.1519/JSC.0b013e318242a32a

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

le Gall, F., Carling, C., Williams, M., & Reilly, T. (2010). Anthropometric and fitness characteristics of international, professional and amateur male graduate soccer players from an elite youth academy. *Journal of Science and Medicine in Sport*, *13*(1), 90–95.

https://doi.org/10.1016/j.jsams.2008.07.004

Malina, R. M., Cumming, S. P., Kontos, A. P., Eisenmann, J. C., Ribeiro, B., & Aroso, J. (2005). Maturity-associated variation in sport-specific skills of youth soccer players aged 13-15 years. *Journal of Sports Sciences*, *23*(5), 515–522. https://doi.org/10.1080/02640410410001729928

Meylan, C., Cronin, J., Oliver, J., & Hughes, M. (2010). Talent Identification in Soccer: The Role of Maturity Status on Physical, Physiological and Technical Characteristics. *International Journal of Sports Science & Coaching*, *5*(4), 571–592. https://doi.org/10.1260/1747-9541.5.4.571

Mirwald, R. L., Baxter-Jones, A. D. G., Bailey, D. A., & Beunen, G. P. (2002). An assessment of maturity from anthropometric measurements. *Medicine & Science in Sports & Exercise*, *34*(4), 689–694. https://doi.org/10.1097/00005768-200204000-00020

Morgans, R., Orme, P., Anderson, L., & Drust, B. (2014). Principles and practices of training for soccer. *Journal of Sport and Health Science*, *3*(4), 251–257. https://doi.org/10.1016/j.jshs.2014.07.002

Murr, D., Feichtinger, P., Larkin, P., O’Connor, D., & Höner, O. (2018). Psychological talent predictors in youth soccer: A systematic review of the prognostic relevance of psychomotor, perceptual-cognitive and personality-related factors. *PLoS ONE*, *13*(10). https://doi.org/10.1371/journal.pone.0205337

Murr, D., Raabe, J., & Höner, O. (2018). The prognostic value of physiological and physical characteristics in youth soccer: A systematic review. *European Journal of Sport Science*, *18*(1), 62–74. https://doi.org/10.1080/17461391.2017.1386719

Murtagh, C. F., Brownlee, T. E., O'Boyle, A., Morgans, R., Drust, B., & Erskine, R. M. (2018). Importance of Speed and Power in Elite Youth Soccer Depends on Maturation Status. *Journal of Strength and Conditioning Research*, *32*(2), 297–303. https://doi.org/10.1519/JSC.0000000000002367

Musculus, L., & Lobinger, B. H. (2018). Psychological characteristics in talented soccer players - Recommendations on how to improve coaches' assessment. *Frontiers in Psychology*, *9*(41). https://doi.org/10.3389/fpsyg.2018.00041

Paul, D. J., & Nassis, G. P. (2015). Physical Fitness Testing in Youth Soccer: Issues and Considerations Regarding Reliability, Validity, and Sensitivity. *Pediatric Exercise Science*, *27*(3), 301–313. https://doi.org/10.1123/mc.2014-0085

Philippaerts, R. M., Vaeyens, R., Janssens, M., Van Renterghem, B., Matthys, D., Craen, R., … Malina, R. M. (2006). The relationship between peak height velocity and physical performance in youth soccer players. *Journal of Sports Sciences*, *24*(3), 221–230. https://doi.org/10.1080/02640410500189371

Pyne, D. B., Spencer, M., & Mujika, I. (2014). Improving the value of fitness testing for football. *International Journal of Sports Physiology and Performance*, *9*(3), 511–514. https://doi.org/10.1123/IJSPP.2013-0453

Rebelo, A., Brito, J., Maia, J., Coelho-e-Silva, M. J., Figueiredo, A. J., Bangsbo, J., … Seabra, A. (2013). Anthropometric Characteristics , Physical Fitness and Technical Performance of Under-19 Soccer Players by Competitive Level and Field Position. *International Journal of Sports Medicine*, *34*(4), 312–317.

Reeves, M. J., Enright, K. J., Dowling, J., & Roberts, S. J. (2018). Stakeholders' understanding and perceptions of bio-banding in junior-elite football training. *Soccer and Society*, *19*(8), 1166–1182. https://doi.org/10.1080/14660970.2018.1432384

Reeves, M. J., McRobert, A. P., Lewis, C. J., & Roberts, S. J. (2019). A case study of the use of verbal reports for talent identification purposes in soccer: A Messi affair! *PLoS ONE*, 1–17.

Reeves, M. J., Roberts, S. J., McRobert, A. P., & Littlewood, M. A. (2018). Factors affecting the identification of talented junior-elite footballers: a case study. *Soccer and Society*, *19*(8), 1106–1121. https://doi.org/10.1080/14660970.2018.1432383

Reilly, T., Williams, A. M., Nevill, A., & Franks, A. (2000). A multidisciplinary approach to talent identification in soccer. *Journal of Sports Sciences*, *18*(9), 695–702.

Roberts, S. J., McRobert, A. P., Lewis, C. J., & Reeves, M. J. (2019). Establishing consensus of position-specific predictors for elite youth soccer in England. *Science and Medicine in Football*, *3*(3), 205–213. https://doi.org/10.1080/24733938.2019.1581369

Saward, C., Morris, J. G., Nevill, M. E., Nevill, A. M., & Sunderland, C. (2016). Longitudinal development of match-running performance in elite male youth soccer players. *Scandinavian Journal of Medicine and Science in Sports*, *26*(8), 933–942. https://doi.org/10.1111/sms.12534

Sieghartsleitner, R., Zuber, C., Zibung, M., & Conzelmann, A. (2019). Science or Coaches ’ Eye ? – Both! Beneficial Collaboration of Multidimensional Measurements and Coach Assessments for Efficient Talent Selection in Elite Youth Football. *Journal of Sports Science and Medicine*, *18*(1), 32–43.

Stølen, T., Chamari, K., Castagna, C., & Wisloff, U. (2005). Physiology of Soccer: An Update.

Sports Medicine, *35*(6), 501–536. https://doi.org/https://doi.org/10.2165/00007256-200535060-00004

Unnithan, V., White, J., Georgiou, A., Iga, J., & Drust, B. (2012). Talent identification in youth soccer. *Journal of Sports Sciences*, *30*(15), 1719–1726. https://doi.org/10.1080/02640414.2012.731515

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Williams, A. M., & Reilly, T. (2000). Talent identification and development in soccer. *Journal of Sports Sciences*, *18*(9), 657–667.

Yin, R. K. (2009). *Case Study Research: Design and Methods* (Fourth Edi). SAGE.

Table 1. Descriptive statistics of raw data from measured variables for coach's subjective ratings of players' and corresponding objective physical performance.

		Coach's Subjective Rating				
		1 Poor	2 Below Average	3 Average	4 Above Average	5 Excellent
YYIRT L1 (m)	Lead	1387 ± 167 (n = 3)	1213 ± 551 (n = 16)	1374 ± 566 (n = 29)	1855 ± 577 (n = 24)	2234 ± 621 (n = 8)
	Assistant	920 ± 396 (n = 3)	1184 ± 409 (n = 5)	1613 ± 501 (n = 22)	1667 ± 711 (n = 41)	1329 ± 615 (n = 9)
CMJ (cm)	Lead	40.4 ± 5.2 (n = 3)	40.7 ± 5.7 (n = 14)	42.2 ± 7.7 (n = 33)	45.9 ± 7.1 (n = 23)	48.9 ± 5.6 (n = 7)
	Assistant	42.3 ± N/A (n = 1)	39.3 ± 3.7 (n = 10)	41.9 ± 7.2 (n = 33)	45.6 ± 7.3 (n = 24)	46.4 ± 7.8 (n = 12)
FMS (score)	Lead	16.3 ± 2.1 (n = 4)	15.8 ± 2.7 (n = 16)	17.0 ± 1.9 (n = 34)	17.2 ± 2.5 (n = 21)	17.6 ± 0.9 (n = 5)
	Assistant	15.5 ± 2.1 (n = 3)	16.5 ± 2.4 (n = 12)	17.3 ± 2.2 (n = 24)	16.5 ± 2.6 (n = 27)	16.9 ± 1.5 (n = 14)

Table 1. *Cont.*

		Coach's Subjective Rating				
		1 Poor	2 Below Average	3 Average	4 Above Average	5 Excellent
Maturity offset (years)	Lead	-1.9 ± 1.6 (n = 7)	-2.4 ± 0.8 (n = 13)	-2.4 ± 1.1 (n = 30)	-1.8 ± 1.5 (n = 22)	-1.8 ± 1.3 (n = 8)
	Assistant	-1.3 ± 2.6 (n = 3)	-2.4 ± 1.0 (n = 6)	-2.3 ± 1.1 (n = 34)	-1.9 ± 1.4 (n = 18)	-2.2 ± 1.3 (n = 19)
5m sprint (s)	Lead	1.14 ± 0.05 (n = 7)	1.06 ± 0.11 (n = 10)	1.06 ± 0.08 (n = 36)	1.03 ± 0.08 (n = 22)	0.94 ± 0.07 (n = 4)
	Assistant	N/A	1.09 ± 0.06 (n = 14)	1.05 ± 0.10 (n = 34)	1.03 ± 0.08 (n = 27)	1.02 ± 0.11 (n = 5)
20 sprint (s)	Lead	3.50 ± 0.15 (n = 7)	3.30 ± 0.29 (n = 10)	3.34 ± 0.19 (n = 36)	3.18 ± 0.21 (n = 22)	3.01 ± 0.17 (n = 5)
	Assistant	3.31 ± 0.02 (n = 3)	3.45 ± 0.13 (n = 7)	3.33 ± 0.26 (n = 28)	3.24 ± 0.21 (n = 35)	3.21 ± 0.25 (n = 7)

Table 2. A Bayesian estimation of the coefficient of variation (R^2) with 95% credible intervals for each of the Bayesian monotonic ordinal regression models and Sklar's ω for agreement.

		Endurance	Power	Movement Quality	Physical Development	Acceleration	Sprint Speed
Lead Coach	R^2	0.23	0.11	0.05	0.03	0.17	0.2
	95% CI	0.08-0.37	0.01-0.23	0.00-0.16	0.00-0.12	0.04-0.32	0.06-0.33
Assistant Coach	R^2	0.03	0.09	0.01	0.02	0.06	0.07
	95% CI	0.00-0.11	0.00-0.22	0.00-0.07	0.00-0.08	0.00-0.18	0.00-0.19
Agreement	Sklar's ω	0.48	0.68	0.49	0.54	0.62	0.62
	Interpretation	Moderate	Substantial	Moderate	Moderate	Substantial	Substantial

Figure captions

Figure 1. Raw data boxplots for lead and assistant coach ratings for: A) Yo-Yo test distance; B) CMJ height; C) FMS score; D) maturity offset years; E) 5m sprint times, and; F) 20m sprint times.

Figure 2. Marginal effects of the predictive Bayesian monotonic ordinal regression models (±95%CI) for lead and assistant coach ratings at population level for: A) Yo-Yo test distance; B) CMJ height; C) FMS score; D) maturity offset years; E) 5m sprint times, and; F) 20m sprint times.