Match performance variables influential in selection into the English national football team

ABSTRACT

Background: In recent years, talent identification has become an imperative tool in football, given the highly lucrative nature of the modern-day game. However, despite attempts to utilize multidimensional assessments for talent identification, there is little evidence to propose a set of factors that can reliably predict future footballing success. **Purpose:** To examine match actions of players in the four major playing positions in football (goalkeepers, defenders, midfielders, forwards) to see if they can distinguish between players selected or not selected for the national team. **Method:** Match actions data of English players playing in the English Premier League (EPL) from 2012-2018 were collected, with analysis done on 828 season-long observations. Discriminant analysis was run to identify performance variables that best distinguished those selected or not selected for the English national team for each position. The discriminant functions were tested on their ability to successfully classify the players into the national team or non-national team categories using the leave-one-out method. **Results:** Four unique sets of critical performance variables were identified using discriminant analysis. Multiple critical performance variables were identified for each position but the performance variable best able to distinguish the two selection groups were: saves made (goalkeepers), clean sheets kept (defenders), tackles made (midfielders), and shots on target (forwards). The four discriminant functions provided high overall percentages of successful classification (>77%). **Conclusion:** The usage of match performance variables profile analysis is a viable addition to existing multidimensional methods of football talent identification.

Keywords: football talent identification, match performance analysis, performance variables

Match performance variables influential in selection into the English national football team

Introduction

Background

In the modern era of sport, footballing success at the elite level involves winning both on and off the pitch. Football has become a highly profitable global brand in the past decade, spurring football clubs to constantly compete against each other to sign the best players and give their team the best chance of winning (Littlewood, Mullen, & Richardson, 2011). Since 2010, football clubs in the English Premier League (EPL) have been spending larger amounts of money on player transfers (Scott & Tweedale, 2018), culminating in a lofty 1.6 billion pounds spent in the 2016 season (Barnard, Dwyer, Wilson, & Winn, 2018). However, these massive investments are often unfounded gambles as clubs that have spent the most at the start of the season have not consistently won the EPL (Scott & Tweedale, 2018).

Beyond the club level, international competitions such as the FIFA World Cup present another set of challenges for coaches attempting to assemble the best team. National team coaches are free to select any players that they want as long as they hold the necessary citizenship (Hall, 2012). However, professional players are involved with their domestic club teams for the majority of each season which limits the interaction between national team coaches and the players that they can select. Furthermore, globalization of modern football has seen an increase in the number of players migrating to play in foreign leagues (Littlewood et al., 2011), which further complicates the scouting and selection process. These circumstances present the need for a set of performance variables which coaches can use to make more informed decisions in player selection. Despite the widespread usage of performance variables in sports to identify the best players among elite athletes, existing literature has been unable to agree on a set of variables that can reliably predict future sporting success (Johnston, Wattie, Schorer, & Baker, 2018).

Research gap and objectives

Given the benefits of being able to correctly identify talented athletes, there has been an abundance of literature in athlete talent identification. Till date, existing methods of identifying talented athletes across multiple sports have been primarily concerned with performance variables relating to physical profiles of athletes and have often been limited to examining athletes 10 to 20 years of age (Johnston et al., 2018). Moreover, despite widespread usage of performance variables to identify talent, current methods have been unable to establish an apparent set of variables that can reliably predict sporting success in the future (Johnston et al., 2018).

Existing football specific talent identification methods commonly examine anthropometric, physiological, psychological and football specific motor skills to distinguish between youth players of different skill groups (HÖner, Votteler, Schmid, Schultz, & Roth, 2015; Reilly, Williams, Nevill, & Franks, 2000). Till date, there have been a number of studies that have proposed multiple sets of key technical skills required by players to succeed in different playing positions (Wiemeyer, 2003). However, as these publications are often based on the personal opinion of coaches, there has not been a large amount of agreeability in key technical football skills that can predict future footballing success (Hughes & Bartlett, 2002). Furthermore, the assessment of youth players on traditional metrics such as anthropometric and physical profiles are problematic due to maturation and training effects, as the characteristics that are innate or evident in pre-adolescence do not always translate successfully when athletes transition into the senior level (Vaeyens, Lenoir, Williams, & Philippaerts, 2008).

Currently, there has been no study examining football specific actions performed during actual professional matches that can distinguish the best players among elite senior football players. Additionally, current literature has not attempted to highlight differences in actual match performances between players selected and not selected for the national team.

Therefore, if a set of crucial match performance variables can be identified for each position, there may be significant benefits for national team coaches particularly in helping them identify potential prospects from a vast pool of available players. Hence, the objective of the current study is to analyze match actions performed during actual matches as a form of match performance variables. The match performance variables of English players playing in the English Premier League will be analyzed to identify a critical set of match performance variables for each position that best distinguishes players selected or not selected for the national team.

As such, the current study seeks to answer the following research questions:

1. What are the critical match performance variables performed by players in each of the four major positions in football during domestic club matches that best distinguishes them from being selected or not selected for the national football team?
2. What is the ability of the set of critical performance variables identified in successfully classifying the players into their respective selection groups?

Significance of study

It was reported by Hunter (2018) that in the first four weeks of a 38-week season, English players accounted for 24,076 minutes played in the EPL. Therefore, if a set of critical performance variables can be identified for each of the four major positions, national team coaches may benefit by using potential findings to guide the allocation of time and financial resources in scouting and recruiting prospective players. Given the uncertainty of current methods in identifying footballing talent, the potential findings may help to augment current approaches of football talent identification by adding an evidence-based approach to recruit players with greater certainty. Additionally, the results may be useful for youth football coaches by placing a greater emphasis on development of skills identified to be valued in senior elite football.

Literature Review

Talent identification in elite football athletes

Given the popularity of football and financial benefits of being able to identify footballing talent among youth players, there has been a large amount of literature on the different approaches towards identifying young footballing talent. Table 1 provides a summary of the literature regarding existing approaches to football talent identification. It is crucial to note that majority of existing literature examined players at the junior level (Johnston et al., 2018). This may be because early identification of talented young players allows them to receive specialized coaching at a young age (Reilly et al., 2000). Additionally, football clubs may receive greater dividends by focusing on youth scouting. Specifically, the Union of European Football Associations (UEFA) has exempted youth player related expenses incurred by football clubs from financial fair play regulations, in an effort to encourage long term investments on youth ("UEFA Club Licensing and Financial Fair Play Regulations," 2015).

Current literature has taken a holistic approach toward talent identification in football, examining factors such as perceptual cognitive awareness in games, sports specific performance measures such as 5-0-5 agility and 40 meter sprints, sport specific skills such as shooting or passing, match activity profiles established from global positioning system (GPS) metrics, and anthropometric measures (Wilson, 2017). However, current metrics assessing youth footballing talent are typically single measures employed in cross sectional studies, and few studies have explored the effectiveness of such talent identification approaches when youth athletes transition into the senior level (Wilson, 2017). It has been proposed that by monitoring athletes over a protracted period of time it is possible to establish the key changes that happen due to development and practice, and ascertain how these changes contribute to the transition toward expert performance at the senior level (Reilly et al., 2000).

Table 1. Summary of current literature in identifying footballing talent.

Study	Methods				Findings
	Research design	Participants	Intervention/Test items	Main Topic	
HÖner, Votteler, Schmid, Schultz, & Roth (2015)	Repeated cross-sectional study Players from youth academy and competence centre completed a heterogeneous test battery.	68,158M highly ranked players within their age group from 12-15 years old	Test battery items: - Sprint, agility, dribbling, ball control, shooting, juggling	- Football specific skills - Speed ability	High internal consistency, test-test reliability and differential stability for overall scores and speed related tests. Overall score, dribbling and juggling scores were the best in differentiating between players of different performance levels and displayed highest criterion-related validity.
Reilly, Williams, Nevill & Franks (2010)	Cross-sectional study Multidisciplinary test battery to youth elite and sub-elite players	31M players aged 15-16 years, 16 elite and 15 sub-elite players	Test battery items: - Somatotype, body composition, body size - Speed, endurance, agility - Technical skill - Anticipation - Task & ego orientation, anxiety	- Anthropometric - Physiological - Psychological - Football specific skills - Speed ability	Elite players had significantly lower body fat percentages, able to exert more aerobic power and more resistant to fatigue. Performance variables best able to discriminate between the groups were agility, sprint time, ego orientation and anticipation skill.
O'Connor, Larkin & Williams (2016)	Cross-sectional study Elite youth male football players assessed on perceptual-cognitive skills and player history variables.	127M elite youth male football plyers	Test battery items: - Adapted version of PHV - Match-play assessment - Player history such as match play, coach-led practice, individual practice hours per month - Decision-making, anticipation, situational probability and pattern recognition	- Perceptual cognitive skills - Match-play analysis - Athlete history	Players offered a scholarship programme exhibited significantly better performances on the perceptual-cognitive skills test. Discriminant function analysis highlighted that recent match play performance, region, number of sports participated in, and combined perceptual-cognitive performance were best able to discriminate the players selection for the scholarship.
Keller, Raynor, Bruce & Iredale (2016)	Cross-sectional study Australian National elite, State elite and Sub-elite players assessed on their technical ability	62M youth football players aged 17.0 ± 0.61 years 18 elite, 22 state elite, 22 sub-elite players	Technical ability assessed by - LSPT - LLPT - Shooting test - Speed dribbling	- Football specific skills	Technical ability tests were able to distinguish youth football players of different playing levels. Scores of three groups differed significantly on the LPST. Sub-elite cohort performed more poorly on the LPT compared to both national and state-elite cohorts. Shooting accuracy and velocity on the dominant foot was able to distinguish the national and sub-elite cohorts, while shooting velocity on the nondominant foot was faster for the national elite players compared to the other cohorts.
Goto, Morris & Nevill (2015)	Cross-sectional study Global Position System devices were used with elite youth football players to establish the match profile of players.	34M English Premier League Academy players, 22 Under 9 and 12 Under 10 players	Match activity profiles examined - Total distance covered - Distance covered at difference speed zones (walking, jogging, low speed, moderate speed, high speed running)	- Physical match activity profile	Players subsequently retained by their respective clubs for future grooming covered a greater distance during six-a-side matches and also a greater distance at low-speed in both absolute and relative terms.

Note: PHV – Participation History Questionnaire, LSPT – Loughborough Short Passing Test, LLPT – Loughborough Long Passing Test

Match performance variables for talent identification

Given the dynamic nature of football, successful performances are an amalgamation of physical, technical and tactical actions from all participating players on the pitch (Bradley et al., 2011). Perhaps due to greater complexity in recording of technical actions performed by players in matches, current literature examining match performances of football players have primarily focused on physical performance metrics, or a mix of physical performance metrics combined with certain technical action metrics (Rampinini, Impellizzeri, Castagna, Coutts, & Wisløff, 2009; Russell, Rees, & Kingsley, 2013). However, it has been proposed that technical actions in football matches may be a better measure of future footballing success (Bush, Barnes, Archer, Hogg, & Bradley, 2015; Lago-Peñas, Lago-Ballesteros, & Rey, 2011; Rampinini et al., 2009; Russell et al., 2013).

In recent years, a number of studies have utilized datasets provided by Opta Sportsdata to examine technical actions exhibited during professional football matches. These studies have attempted to identify differences in technical actions exhibited between players in different teams (Liu, Gómez, Gonçalves, & Sampaio, 2016), players in similar positions on different teams (Sainz de Baranda et al., 2019) and also players in different positions (Yi, Jia, Liu, & Gómez, 2018). The results of such studies suggest that despite playing in similar positions, players may have distinguishable match performance profiles. For example, Sainz de Baranda et al. (2019) found that female elite goalkeepers that played for higher performing teams executed higher number of offensive technical actions and greater number of successful passes in different areas of the field. Furthermore, Liu et al. (2016) found that technical action match profiles of players in strong and weak teams were significantly different across different positions. These findings suggest that technical actions executed during matches may be used to build a match profile that could be employed for football talent identification.

Discriminant analysis in sports and common usages

As highlighted in the previous section, analyzing match performance variable profiles involves simultaneously comparing the differences of several performance variables performed by two or more groups. Such statistical analyses can be conducted through the discriminant analysis method which involves simultaneously examining a group of continuous variables to determine which are the variables that best discriminate between two or more naturally occurring groups (Klecka, 1980).

For example, existing literature has utilized discriminant analysis methods on football match statistics in order to identify statistics that best discriminate between the performance of teams. Lago-Peñas et al. (2011) found that match statistics best able to discriminate between winning and losing teams in the UEFA Champions League were the number of shots on target, crosses, ball possession, whether the match was played at home or away, and quality of the opposition. Through discriminant analysis, certain studies have also highlighted characteristics that best discriminate between the playing level of players. In a study applying a holistic test battery on 31 elite and sub-elite youth players, Reilly et al. (2000) proposed that measures of agility, sprint time, ego orientation and anticipation skill performed best in distinguishing between the elite and sub-elite players. The above-mentioned studies propose the value of using a discriminant analysis to highlight performance variables that best differentiate players selected or not selected for the national team.

Method

Research design

The current study employed a retrospective study design. The independent variables used in the discriminant analysis were the match performance variables of the players while the categorical dependent variable was the national team selection status of the players.

Participants

The players included in the analysis were English players playing in the EPL between 2012 to 2018. The players were categorized into four major playing positions based on their categorization on the EPL website. Match performance variable data of 401 English players were collected and categorized into the four major positions: goalkeepers (N = 26 players, n = 63 season-long observations), defenders (N = 141 players, n = 371 season-long observations), midfielders (N = 150, n = 429 season-long observations), and forwards (N = 84, n = 205 season-long observations).

Match performance variables

Season-long observations of match performance variables and total playing time of the players were collected from the Opta Sports database hosted on the official EPL website (www.premierleague.com/stats). Opta Sports is responsible for collecting and analyzing official EPL performance data, and all data collected has been checked through to ensure accuracy (Greig, 2017). The tracking system employed by OPTA Sports to collect data on the individual match actions of football players has been found to have an acceptable inter-operator reliability, with standardized typical error varying from 0.00 to 0.37 and intra-class correlation coefficients ranging from 0.88 to 1.00 (Liu, Hopkins, Gómez, & Molinuevo, 2013).

A total of 2280 matches were played in the EPL from 2012 to 2018 and there were a total of 1092 available season-long observations across the six seasons. As the count values of the performance variables for each player are related to the duration each player spends on the pitch in each season, it was necessary to normalize the performance variable counts for a fair comparison. Norusis (1993) proposes that using derived-rate variables, in which original season-long performance variable statistics are divided by the time that the player played throughout the season, makes the discriminant analysis more robust. Using derived-rate variables with discriminant analysis is particularly relevant to the current study as it has been previously used to identify performance variables that best discriminate between players of different playing levels but playing in the same position (Sampaio, Janeira, Ibáñez, & Lorenzo, 2006). Similarly, season-long match performance variable observations for each player were divided by the minutes they played throughout the respective EPL season. Performance variables such as number of clean sheets kept and percentage-based variables (shot accuracy, tackle success and cross accuracy percentage) were not normalized due to nature of the data.

Table 2 details performance variables that were used in the final analysis for each player position and the established operational definitions of the Opta Sports performance variables (Liu et al., 2013; OPTA, 2012) collected and analyzed in this study.

Table 2. Operational definitions of performance variables collected for each player position.

Performance Variable	Definition	Player position in which performance variable was included in analysis
Saves	A goalkeeper preventing the ball from entering the goal with any part of his body when facing an intentional attempt from an opposition player	GK
Punches	A high ball that is punched clear by the goalkeeper. The keeper must have a clenched fist and attempting to clear the high ball rather than claim it.	GK
High claim	A high ball played from a cross that is caught by the goalkeeper.	GK

Table 2 (continued).

Sweep clearance	A keeper sweeper is given anytime a goalkeeper anticipates danger and rushes off their line to try to either cut out an attacking pass (in a race with the opposition player) or to close-down an opposition player.	GK
Clean sheets	A player or team who does not concede a goal for the full match.	GK, DF
Throws	An intentionally played ball from the keeper to another player using the hands.	GK
Catches	A high ball played into the penalty area that is caught by the goalkeeper.	GK
Penalty saves	A goalkeeper preventing the ball from entering the goal from a penalty.	GK
Fouls	A foul conceded is defined as any infringement that is penalised as foul play by a referee.	GK, DF, MD,
Errors leading to goal	When a player makes an error, which leads to a goal conceded	GK, DF, MD,
Tackles	A tackle is defined as where a player connects with the ball in a ground challenge where he successfully takes the ball away from the player in possession.	DF, MD, FW.
Blocks	This is where a player blocks a shot on target from an opposing player.	DF, MD,
Clearances	This is a defensive action where a player kicks the ball away from his own goal with no intended recipient.	DF, MD,
Headed clearances	This is a defensive action where a player heads the ball away from his own goal with no intended recipient.	FW
Interceptions	This is where a player reads an opponent's pass and intercepts the ball by moving into the line of the intended pass.	DF, MD, FW.
Recovery	This is where a player recovers the ball in a situation where neither team has possession or where the ball has been played directly to him by an opponent, thus securing possession for their team.	MD
Duels won	This is where a player wins a 50-50 contest between two players of opposing sides in the match.	DF, MD,
Duels lost	This is where a player loses a 50-50 contest between two players of opposing sides in the match.	DF,
Aerial challenge won	This is where two players challenge in the air against each other. The player that wins the ball is deemed to have won the duel.	MD
Aerial challenge lost	This is where two players challenge in the air against each other. The player that loses the ball is deemed to have lost the duel.	MD
Tackle success percentage	Percentage of successful tackles out of the total attempted	DF, MD,
Goals	Goal scored	DF, MD
Goals with head	Goal scored with the head	FW
Goals with right foot	Goal scored with the right foot	FW
Goals with left foot	Goal scored with the left foot	FW

Table 2 (continued).

Shots on target	A shot on target is defined as any goal attempt that (1) goes into the net regardless of intent, (2) is a clear attempt to score that would have gone into the net but for being saved by the goalkeeper or is stopped by a player who is the last-man with the goalkeeper having no chance of preventing the goal (last line block).	MD, FW
Shot accuracy	A calculation of shots on target divided by all shots (excluding blocked attempts and own goals).	FW
Freekicks scored	Goal scored from a freekick.	MD, FW
Penalties scored	Goal scored from a penalty kick.	FW
Assists	This is where the player plays final pass or cross leading to the recipient of the ball scoring a goal.	DF, FW
Crosses	Any intentional played ball from a wide position intending to reach a team mate in a specific area in front of the goal.	DF, FW
Cross accuracy	Percentage of successful crosses out of the total attempted.	MD
Through balls	A pass splitting the defence for a team-mate to run on to.	DF, MD
Big chances created	Creation of a situation where a player should reasonably be expected to score, usually in a one on one scenario or from very close range when the ball has a clear path to goal and there is low to moderate pressure on the shooter.	DF, MD, FW
Offsides	Awarded to the player deemed to be in an offside position where a free kick is awarded.	DF

Note. GK – Goalkeepers, DF – Defenders, MD – Midfielders, FW – Forwards

Selection for England men's national football team

International match representation data of the players for each season were collected in order to classify the season-long observations of the players into the selected for national team (S-NT) and not selected for national team (NS-NT) categories. A breakdown of the players in each category is included in Tables 3 and 4. The England national team representation data from 2012 to 2018 was obtained from www.soccerway.com. The aforementioned website is a football data website that has been widely used in football analytics studies and has been found to be reliable (Matesanz, Holzmayer, Torgler, Schmidt, & Ortega, 2018; Pollard, Armatas, Hojjat, & Sani, 2017). Players were classified as S-NT for that season-long observation if they were on the official team list for any of England's matches for the World Cup, European Championship, World Cup qualification, European Championship qualification, UEFA Nations League or friendly games.

Data collection

To minimize likelihood of human error in collecting data from such a large number of webpages, a web-scraping and database generation script was written in Python, Version 3.7 to automate reading of webpages, obtaining the necessary data and saving the data in Microsoft Excel, Version 16.17, for data analysis. A copy of the Python code can be found in Appendix A.

Statistical analysis

Discriminant analysis was performed on each set of performance variables for each of the four positions to determine which performance variables were most useful in discriminating between players in the S-NT and NS-NT groups. This was done by examining performance variables with structure coefficients greater than |0.30|, as proposed by Tabachnick, Fidell, & Ullman (2007). The interpretation was that performance variables with greater absolute values in their structure coefficients had greater ability to discriminate between groups (Sampaio et al., 2006). Performance variables with structure coefficients less than |0.30| were excluded from the discriminant function.

Based on data accessible on the EPL website, each of the player positions had a unique set of performance variable data available. Oberstone (2010) proposed that there is a considerable redundancy in each set of position specific performance variables collected by OPTA, especially for the goalkeeper position. The redundancy presents the issue of multicollinearity, in which there is an increased correlation between the predictor variables in a discriminant analysis. For example, performance variables such as total goals scored are positively correlated with variables such as goals scored with the right foot (or left foot). Büyüköztürk & Çokluk-Bökeouglu (2008) suggests that predictive power of the discriminant analysis is compromised when there is multicollinearity. The multicollinearity was addressed by trying different specifications of a model using the same data to see if innocuous changes such as adding or dropping a variable produces big shifts, and looking for changes in the signs of effects that seem theoretically questionable when variables are added, as proposed by Williams (2015). Furthermore, discriminant analysis assumes that predictor variables should have a multivariate normal distribution (Klecka, 1980). Shapiro-Wilk test and box-plots revealed minor violations of normality in the performance variables. Nevertheless, this was to be expected as individual performances in sport have been proposed to follow a Paretian power law distribution and not a normal distribution (O'Boyle Jr & Aguinis, 2012). However, Klecka (1980) proposes that discriminant analysis is fairly robust to violations of normality, equality of variances and multicollinearity assumptions but highly sensitive to outliers. Hence, univariate outliers were identified using box-plots and removed. A total of 11 goalkeeper observations, 85 defender observations, 105 midfielder observations and 63 forward observations were removed from analysis.

Lastly, accuracy of the discriminant functions generated were assessed using the leave-one-out method of cross-validation proposed by Norusis (1993) which utilizes subsets of data for training and testing. The method generates the discriminant function on all but one of the participants, who is then tested for classification of group membership based on the discriminant function generated. The process is then repeated for each participant and the correct classification percentage is subsequently generated. The statistical analyses were conducted using IBM SPSS Statistics software version 23.0.0 and the significance level was set at $p \leq 0.05$.

Results

Critical performance variables for each playing position

The mean and standard deviation counts of each performance variable for every 10 minutes played for each of the four playing positions are depicted in Tables 3 and 4.

Table 3. Descriptive statistics of the performance variables for the goalkeeper position

Performance Variable	Goalkeepers	
	S-NT (n = 22)	NS-NT (n = 36)
Saves	0.29 ± 0.16	0.11 ± 0.11
Punches	0.04 ± 0.03	0.02 ± 0.02
High Claim	0.10 ± 0.07	0.04 ± 0.05
Sweep clearance	0.04 ± 0.03	0.02 ± 0.03
Clean sheets*	6.82 ± 4.49	1.92 ± 2.52
Throws	0.32 ± 0.18	0.11 ± 0.11
Catches	0.04 ± 0.03	0.02 ± 0.02
Penalty saves	0.00 ± 0.00	0.00 ± 0.00
Fouls	0.00 ± 0.00	0.00 ± 0.00
Errors leading to goal	0.01 ± 0.01	0.00 ± 0.00

Note. Performance variables denoted with asterisk (*) are not derived-rate variables, values are mean ± S.D. counts per 10 minutes played.

Table 4. Descriptive statistics of the performance variables for the three outfield positions

Performance Variable	Defenders		Midfielders		Forwards	
	S-NT (n = 53)	NS-NT (n = 239)	S-NT (n = 47)	NS-NT (n = 283)	S-NT (n = 22)	NS-NT (n = 126)
Tackles	0.11 ± 0.06	0.08 ± 0.07	0.13 ± 0.07	0.08 ± 0.08	0.05 ± 0.05	0.03 ± 0.03
Blocks	0.01 ± 0.01	0.00 ± 0.01	0.03 ± 0.02	0.01 ± 0.01	-	-
Clearances	0.34 ± 0.23	0.23 ± 0.22	0.07 ± 0.05	0.04 ± 0.05	-	-
Headed clearances	-	-	-	-	0.02 ± 0.02	0.02 ± 0.03
Interceptions	0.11 ± 0.06	0.08 ± 0.07	0.07 ± 0.04	0.05 ± 0.06	0.03 ± 0.03	0.01 ± 0.02
Recovery	-	-	0.38 ± 0.21	0.21 ± 0.21	-	-
Errors leading to goal	0.00 ± 0.00	0.00 ± 0.00	-	-	-	-
Duels won	0.35 ± 0.15	0.22 ± 0.19	0.33 ± 0.16	0.18 ± 0.17	-	-
Duels lost	0.23 ± 0.10	0.16 ± 0.13	-	-	-	-
Aerial challenge won	-	-	0.05 ± 0.03	0.03 ± 0.04	-	-
Aerial challenge lost	-	-	0.07 ± 0.05	0.04 ± 0.04	-	-
Clean sheets*	6.40 ± 3.67	3.08 ± 3.16	-	-	-	-

Table 4 (continued).

Fouls	0.00 ± 0.00	0.00 ± 0.00	0.06 ± 0.04	0.04 ± 0.04	-	-
Tackle success percentage*	0.74 ± 0.09	0.70 ± 0.22	0.74 ± 0.09	0.65 ± 0.29	-	-
Goals	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.01	0.00 ± 0.00	-	-
Goals with head	-	-	-	-	0.00 ± 0.00	0.00 ± 0.00
Goals with right foot	-	-	-	-	0.01 ± 0.01	0.00 ± 0.01
Goals with left foot	-	-	-	-	0.00 ± 0.00	0.00 ± 0.00
Shots on target	-	-	0.03 ± 0.02	0.01 ± 0.02	0.07 ± 0.04	0.03 ± 0.03
Shot accuracy*	-	-	-	-	0.45 ± 0.16	0.29 ± 0.22
Freekicks scored	-	-	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Penalties scored	-	-	-	-	0.00 ± 0.00	0.00 ± 0.00
Assists	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.01	0.00 ± 0.00	0.01 ± 0.01	0.00 ± 0.01
Crosses	0.09 ± 0.10	0.06 ± 0.08	-	-	0.08 ± 0.11	0.03 ± 0.05
Cross Accuracy*	-	-	0.19 ± 0.09	0.15 ± 0.14	-	-
Through balls	0.00 ± 0.00	0.00 ± 0.00	0.02 ± 0.01	0.01 ± 0.01	-	-
Big chances created	0.00 ± 0.01	0.00 ± 0.00	0.01 ± 0.01	0.00 ± 0.01	0.01 ± 0.01	0.00 ± 0.01
Offsides	0.00 ± 0.00	0.00 ± 0.00	-	-	-	-

Note. Performance variables denoted with asterisk (*) are not derived-rate variables, values are mean ± S.D. counts per 10 minutes played.

The discriminant function structure coefficients presented in Tables 5, 6 and Figure 1 describe match performance variables exhibited in EPL matches that best discriminate between English players in the S-NT and NS-NT groups. The discriminant function structure coefficients quantify ability of each performance variable to maximize differences between the match performance variable means of players in the S-NT and NS-NT groups. Larger coefficient magnitudes represent a greater contribution of the performance variable to the discriminant function (Sampaio et al., 2006). Performance variables indicated in the green and red zones in Figure 1 represent performance variables with a structure coefficient greater than |0.30| and were included in the discriminant function for that playing position.

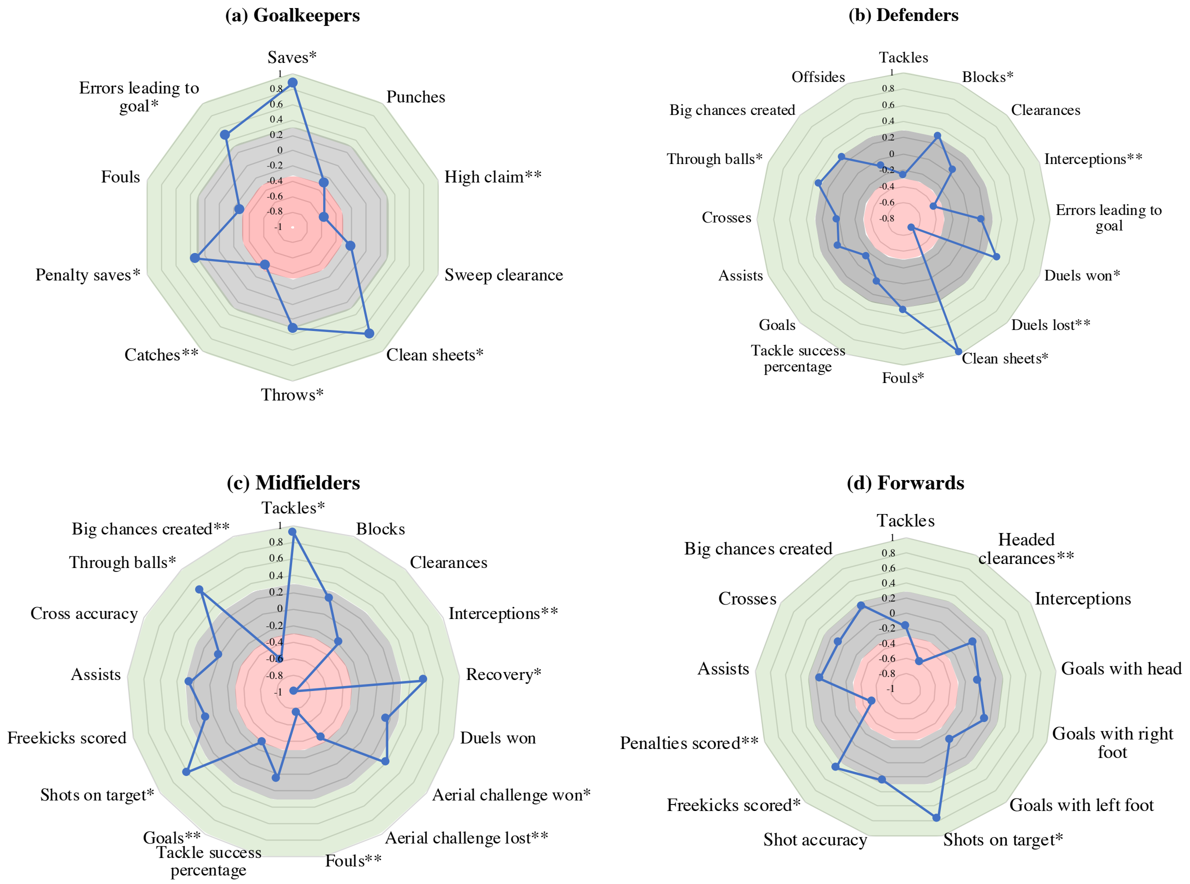

Figure 1. Structure coefficients for performance variables in each position.

The Wilks' Lambda value for each discriminant function was interpreted as how well the discriminant function separated the season-long observations into the two groups, with smaller values indicating a greater discriminatory ability (Klecka, 1980). Additionally, larger eigenvalues indicated greater variance in which selection into the English national team is explained by the discriminant function (Klecka, 1980).

Table 5. Discriminant function structure coefficients for the goalkeeper position and tests of statistical significance.

Performance Variable	Goalkeepers
Saves	.88*
Punches	-.29
High claim	-.56*
Sweep clearance	-.20
Clean sheets[a]	.71*
Throws	.32*
Catches	-.39*
Penalty saves	.33*
Fouls	-.27
Errors leading to goal	.49*
Wilks' Lambda	.508
Chi-square	34.55
P-value	<.001
Eigenvalue	.97
Canonical correlation	.70

Note. Structure coefficients denoted with asterisk (*) indicate value greater than |0.30|.
[a] Indicates performance variables that are not derived-rate variables.

For the goalkeeper position, discriminant analysis revealed an emphasis on saves made (0.88), clean sheets (0.71), errors leading to goal conceded (0.49), penalty saves (0.33), and throws (0.32), and de-emphasis on high claims (-0.56) and catches (-0.39). The Wilks' Lambda value was 0.508 and eigenvalue was 0.97.

Table 6. Discriminant function structure coefficients for the outfielder positions and tests of statistical significance.

Performance Variable	Defenders	Midfielders	Forwards
Tackles	-.26	.91*	-.17
Blocks	.31*	.20	-
Clearances	.05	-.19	-
Headed clearances	-	-	-.60*
Interceptions	-.40*	-.98*	.10
Recovery	-	.58*	-
Errors leading to goal	.15	-	-
Duels won	.44*	.16	-
Duels lost	-.65*	-	-
Aerial challenge won	-	.40*	-
Aerial challenge lost	-	-.36*	-
Clean sheets[a]	.97*		-
Fouls	.32*	-.75*	-

Table 6 (continued).

Tackle success percentage[a]	.04	.05	-
Goals	-.16	-.30*	-
Goals with head	-	-	-.03
Goals with right foot	-	-	.12
Goals with left foot	-	-	-.11
Shots on target	-	.60*	.77*
Shot accuracy[a]	-	-	.25
Freekicks scored	-	.09	.39*
Penalties scored	-	-	-.53*
Assists	.07	.25	.14
Crosses	.02	-	.08
Cross accuracy[a]	-	-.01	-
Through balls	.33*	.65*	-
Big chances created	.27	-.58*	.24
Offsides	-.09	-	-
Wilks' Lambda	.776	.710	.721
Chi-square	71.28	109.57	45.54
P-value	<.001	<.001	<.001
Eigenvalue	.29	.41	.39
Canonical correlation	.47	.54	.53

Note. Structure coefficients denoted with asterisk (*) indicate value greater than |0.30|.
[a] Indicates performance variables that are not derived-rate variables.

For the defender position, analysis revealed an emphasis on clean sheets (0.97), duels won (0.44), through balls played (0.33), fouls (0.32), and blocks made (0.31), and de-emphasis on duels lost (-0.65) and interceptions (-0.40). The Wilks' Lambda value was 0.776 and eigenvalue was 0.29. For the midfielder position, analysis revealed an emphasis on tackles made (0.91), through balls played (0.65), shots on target (0.60), recoveries (0.58), and aerial challenges won (0.40), and de-emphasis on interceptions (-0.98), fouls conceded (-0.75), big chances created (-0.58), aerial challenges lost (-0.36), and goals scored (-0.30). The Wilks' Lambda value was 0.710 and the eigenvalue was 0.41. Lastly, for the forward position, analysis revealed an emphasis on shots on target (0.77) and freekicks scored (0.39), and de-emphasis on headed clearances (-0.60) and penalties scored (-0.53). The Wilks' Lambda value was 0.721 and eigenvalue was 0.39.

Classification ability of each discriminant function

As mentioned earlier, the leave-one-out test describes the ability of the discriminant function to correctly classify players into the S-NT or NS-NT groups. Table 7 summarizes the classification ability of the discriminant function generated for each player position. Altogether, the discriminant function for each position had a successful classification of 84.5% for goalkeepers, 77.7% for defenders, 83.6% for midfielders, and 80.4% for forwards. In particular, the discriminant functions performed well in classifying goalkeepers and midfielders that were not selected for the national team (88.9% and 85.2% respectively).

Table 7. Classification results matrix for the English players' actual and predicted selection or non-selection for the national team based on the discriminant function derived for each playing position

Position	Actual group	Predicted group		
		Combined overall	S-NT	NS-NT
Goalkeepers	S-NT (*n* = 22)	84.5%	77.3%	22.7%
	NS-NT (*n* = 36)		11.1%	88.9%
Defenders	S-NT (*n* = 53)	77.7%	64.2%	35.8%
	NS-NT (*n* = 239)		19.3%	80.7%
Midfielders	S-NT (*n* = 47)	83.6%	74.5%	25.5%
	NS-NT (*n* = 283)		14.8%	85.2%
Forwards	S-NT (*n* = 22)	80.4%	68.2%	31.8%
	NS-NT (*n* = 126)		17.5%	82.5%

Note: Percentage values are generated based on the number of cases classified to each group using the discriminant function derived for each playing position.

Discussion

The purpose of the current study was to examine match actions performed by English players in EPL matches to identify a set of match performance variables for the four major player positions that can distinguish between players selected or not selected for the English national team. Performance variables were collected as season-long observations and were derived from matches played in the EPL from 2012 to 2018. As the performance variables were count data directly related to the time spent playing, performance variables were normalized into derived-rate values. Discriminant analysis was run to identify match performance variables best able to distinguish between players in the national team or non-national team groups. Next, the discriminant functions were assessed on their ability to correctly classify the players into the S-NT and NS-NT groups based on critical match performance variables.

The results of this study present novel insights to the understanding of talent identification in football. Existing research on identifying footballing talent has primarily focused on youth players, during which individuals are still in the process of attaining expertise in their sport (Goto, Morris, & Nevill, 2015; HÖner et al., 2015; Reilly et al., 2000). Our results suggest that using match performance variables as a means of talent identification is viable at the senior level, and that elite football players can be classified on their potential selection for the national team based on a set of critical performance variables unique to each position. The sets of performance variables identified had high ability (>77%) in correctly classifying players into their national team selection status. According to Sampaio et al. (2006), this indicates the quality of the discriminant functions and power of the structure coefficients in explaining the variability between the two groups. The ability to distinguish talent in senior elite football players is particularly important as: a) individual differences in physical performances are generally attenuated in young adulthood (Malina et al., 2005), b) talented junior elite athletes often do not successfully transition into the senior level due to a over reliance on strength and size advantages they acquire from maturing earlier than their peers (Vaeyens, Philippaerts, & Malina, 2005). As such, our findings present an alternative method of talent identification by examining athletes when they are at the senior level and naturally occurring differences in physical strength and size are minimized. Examining athletes at the senior level addresses certain concerns about early talent identification such as a bias toward early maturing individuals. This is particularly important as physical superiority experienced in youth does not reliably translate to success at the senior level (le Gall, Carling, Williams, & Reilly, 2010; Williams & Reilly, 2000).

Additionally, by identifying the set of performance variables that can successfully distinguish talented players at the elite level, coaches and football team owners can make better informed decisions during player selection through adding match performance variable profile analysis to their selection process. It has been highlighted that combining coach assessments with scientific, multidimensional assessment tools may help to negate mutual weaknesses of each approach when identifying footballing talent (Sieghartsleitner, Zuber, Zibung, & Conzelmann, 2019). The critical performance variables identified in the current study are consistent with existing literature in football performance analysis. For example, an existing review highlighted that saves made, and throws played were key technical performance indicators for goalkeepers (Hughes et al., 2012), similar to findings of the current study. Furthermore, Liu et al. (2016) found that players in the top three teams of the Spanish La Liga had more shots on target and through balls played, but fewer clearances executed compared to players from bottom three teams. Furthermore, Rampinini et al. (2009) found that tackles made and shots on target performed best in differentiating top and bottom teams in the Italian Serie A. Similarly, the current study highlighted an emphasis on performance variables such as shots on target (midfielders and forwards), through balls played (defenders and midfielders), tackles made (midfielders) and a de-emphasis on headed clearances (forwards) which were best able to discriminate between players in each position.

However, both aforementioned studies identified short passes and assists as performance variables differentiating top and bottom teams, with top teams achieving more successful passes and assists (Liu et al., 2016; Rampinini et al., 2009). This suggests that although there appears to be certain universal, core performance variables that influence footballing success, there are certain critical performance variables that differ from those identified for English players in the EPL. These differences may be attributed to the difference in playing styles between different football leagues. Sarmento et al. (2013) proposed that playing style of the different leagues, which has a pivotal impact on the match performance variable profile, are greatly influenced by cultural aspects, in addition to strategic-tactical factors and individual skills. A qualitative review of playing styles in the top three European football leagues revealed that the EPL is characterized by a distinctively direct style of play, while La Liga teams focus on maintaining control throughout the game, and teams in the Serie A are characterized by a tactically defensive style (Sarmento et al., 2013). These findings may explain why short passes and successful passes are more valued as the playstyles of the La Liga and Serie A focus on retaining possession of the ball.

A review of literature concerning match analysis in football by Sarmento et al. (2014) found that the physical and technical demands of football players differ based on their playing position on the field. Similarly, the findings of the current study identified a unique set of performance variables critical to each playing position. These findings may be useful in developing training programs for specific positions. For example, football coaches may find it beneficial to place greater emphasis on helping young players develop skills identified to be valued by national team coaches. This is particularly important as it has been highlighted by

Haugaasen & Jordet (2012) that football-specific practice in early developmental phases are critical in achieving expertise in football.

Till date and to the best of knowledge, there is no existing literature that has utilized match actions executed during matches as a means of identifying footballing talent. Furthermore, this is the first study utilizing match actions to examine differences between players selected and not selected for national teams at the senior elite level.

Limitations and future research

A major limitation of the current study was that a number of performance variables available for each position had a high degree of collinearity, which resulted in a number of performance variables being unsuitable for analysis. However, this multicollinearity was to be expected as many actions in football are closely linked. For example, although the actions of tackling and making passes are not directly related to each other, players who make more tackles are likely to have more possession of the ball, and hence may be likely to make more passes.

Furthermore, the current study had to remove a significant number of data cases containing outliers as discriminant analysis is particularly sensitive to outliers. The presence of outliers in the current study is consistent with prior studies investigating technical and physical parameters in the EPL, which have also identified outliers (Barnes, Archer, Hogg, Bush, & Bradley, 2014). However, Aitken (2004) proposes that data distributions of elite performers are frequently positively skewed, and that lognormal distributions may provide a better model. Hence, when identifying outliers in future analyses utilizing match actions as performance variables, consideration should be taken that frequency distribution of such discrete events, such as shots on target or tackles made, do not follow a normal distribution (Nevill, Atkinson, Hughes, & Cooper, 2002).

In the seven years that the dataset was drawn from, the England national team has had three different managers. Till date, there have not been studies investigating the effect of management changes on the players selected for the national team. Hence, it is unknown if changes in national team managers affects the type of players selected for the national team and if there are any subsequent effects on the critical match performance variable profiles. Future research may benefit from running the same analysis for each duration of the national team manager's tenure to see if there were any differences in players selected for the national team.

Lastly, English players playing in the EPL were chosen for analysis in the current study as majority of players chosen for the national team are playing in the EPL. This was done to limit the effect that different football leagues have on playing styles, which might have been a confounding factor (Sarmento et al., 2013). However, this compromises the external validity of the sets of critical performance variables identified for each position as different playing styles found in the different professional leagues may have an effect on the match performance variable profiles. This can be addressed in future research by drawing data from different professional football leagues and different national teams.

Conclusion

The results from this study suggest that utilizing a set of critical match performance variables unique to playing position, to assess elite English players at the senior level, has a good ability (>77%) to successfully determine their selection into the English national team. Although future work should address external validity concerns and employ a more diverse dataset to improve generalizability of applying the four sets of critical match performance variables to a global football setting, findings of the current study present the viability of adding match performance variable profile analysis to existing holistic, multidimensional methods of identifying footballing talent.

Word Count: 4971 words

References

Aitken, T. (2004). Statistical analysis of top performers in sport with emphasis on the relevance of outliers. *Sports Engineering*, *7*(2), 75–88. https://doi.org/10.1007/BF02915919

Barnard, M., Dwyer, M., Wilson, J., & Winn, C. (2018). Annual Review of Football Finance 2018 Sports Business Group, (June), 36. Retrieved from www.deloitte.co.uk/sportsbusinessgroup

Barnes, C., Archer, D. T., Hogg, B., Bush, M., & Bradley, P. S. (2014). The evolution of physical and technical performance parameters in the English Premier League. *International Journal of Sports Medicine*, *35*(13), 1095–1100. https://doi.org/10.1055/s-0034-1375695

Bradley, P. S., Carling, C., Archer, D., Roberts, J., Dodds, A., Di Mascio, M., … Krustrup, P. (2011). The effect of playing formation on high-intensity running and technical profiles in English FA Premier League soccer matches. *Journal of Sports Sciences*, *29*(8), 821–830. https://doi.org/10.1080/02640414.2011.561868

Bush, M., Barnes, C., Archer, D. T., Hogg, B., & Bradley, P. S. (2015). Evolution of match performance parameters for various playing positions in the English Premier League. *Human Movement Science*, *39*, 1–11. https://doi.org/10.1016/J.HUMOV.2014.10.003

Büyüköztürk, S., & Çokluk-Bökeouglu, Ö. (2008). Discriminant function analysis: Concept and application. *Eurasian Journal of Educational Research*, *33*, 73-92.

Goto, H., Morris, J. G., & Nevill, M. E. (2015). Match analysis of U9 and U10 English premier league academy soccer players using a global positioning system: Relevance for talent identification and development. *The Journal of Strength & Conditioning Research*, *29*(4), 954–963. https://doi.org/10.1519/JSC.0b013e3182a0d751

Greig, F. (2017). Dexterity, accuracy and FIFA skills - the people behind Opta’s football data.

Retrieved March 22, 2019, from https://inews.co.uk/sport/football/opta-data-analyst-stats-football/

Hall, C. D. (2012). Fishing for all-stars in a time of global free agency: Understanding FIFA eligibility rules and the impact on the US men's national team. *Marquette Sports Law Review*, *23*(1).

Haugaasen, M., & Jordet, G. (2012). Developing football expertise: A football-specific research review. *International Review of Sport and Exercise Psychology*, *5*(2), 177–201. https://doi.org/10.1080/1750984X.2012.677951

HÖner, O., Votteler, A., Schmid, M., Schultz, F., & Roth, K. (2015). Psychometric properties of the motor diagnostics in the German football talent identification and development programme. *Journal of Sports Sciences*, *33*(2), 145–159. https://doi.org/10.1080/02640414.2014.928416

Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences*, *20*(10), 739–754. https://doi.org/10.1080/026404102320675602

Hughes, M. D., Caudrelier, T., James, N., Redwood-Brown, A., Donnelly, I., Kirkbride, A., & Duschesne, C. (2012). Moneyball and soccer - An analysis of the key performance indicators of elite male soccer players by position. *Journal of Human Sport and Exercise, 7*(2), 402-412. https://doi.org/10.4100/jhse.2012.72.06

Hunter, A. (2018). More talent, fewer minutes: English players suffering in Premier League | Football | The Guardian. Retrieved March 22, 2019, from https://www.theguardian.com/football/2018/sep/05/english-premier-league-gareth-southgate

Johnston, K., Wattie, N., Schorer, J., & Baker, J. (2018). Talent identification in sport: A systematic review. *Sports Medicine*, *48*(1), 97–109. https://doi.org/10.1007/s40279-017-

0803-2

Keller, B. S., Raynor, A. J., Bruce, L., & Iredale, F. (2016). Technical attributes of Australian youth soccer players: Implications for talent identification. *International Journal of Sports Science & Coaching*, *11*(6), 819-824. https://doi.org/10.1177/1747954116676108

Klecka, W. R. (1980). *Discriminant analysis*. Retrieved from https://books.google.com.sg/books/about/Discriminant_Analysis.html?id=9vrgGJy7XOQC&redir_esc=y

Lago-Peñas, C., Lago-Ballesteros, J., & Rey, E. (2011). Differences in performance indicators between winning and losing teams in the UEFA Champions League. *Journal of Human Kinetics*, *27*(1), 135–146. https://doi.org/10.2478/v10078-011-0011-3

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

Littlewood, M., Mullen, C., & Richardson, D. (2011). Football labour migration: An examination of the player recruitment strategies of the "big five" European football leagues 2004-5 to 2008-9. *Soccer and Society*, *12*(6), 788–805. https://doi.org/10.1080/14660970.2011.609680

Liu, H., Gómez, M.-A., Gonçalves, B., & Sampaio, J. (2016). Technical performance and match-to-match variation in elite football teams. https://doi.org/10.1080/02640414.2015.1117121

Liu, H., Hopkins, W., Gómez, A. M., & Molinuevo, S. J. (2013). Inter-operator reliability of live football match statistics from OPTA Sportsdata. *International Journal of Performance Analysis in Sport*, *13*(3), 803–821. https://doi.org/10.1080/24748668.2013.11868690

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

Matesanz, D., Holzmayer, F., Torgler, B., Schmidt, S. L., & Ortega, G. J. (2018). Transfer market activities and sportive performance in European first football leagues: A dynamic network approach. *PLOS ONE*, *13*(12), e0209362. https://doi.org/10.1371/journal.pone.0209362

Nevill, A. M., Atkinson, G., Hughes, M. D., & Cooper, S. M. (2002). Statistical methods for analysing discrete and categorical data recorded in performance analysis. *Journal of Sports Sciences*, *20*(10), 829–844. https://doi.org/10.1080/026404102320675666

Norusis, M. J. (1993). *SPSS: SPSS for Windows, base system user's guide release 6.0*. Retrieved from https://books.google.com.sg/books/about/SPSS_for_Windows.html?id=1QlHAAAAMAAJ&redir_esc=y

O'Boyle Jr, E., & Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology*, *65*(1), 79–119. https://doi.org/10.1111/j.1744-6570.2011.01239.x

O'Connor, D., Larkin, P., & Williams, A. (2016). Talent identification and selection in elite youth football: An Australian context. *European Journal of Sport Science*, *16*(7), 837-844. https://doi.org/10.1080/17461391.2016.1151945

Oberstone, J. (2010). Comparing English premier league goalkeepers: Identifying the pitch actions that differentiate the best from the rest. *Journal of Quantitative Analysis in Sports*, *6*(1). https://doi.org/10.2202/1559-0410.1221

OPTA. (2012). Definitions OPTA and pack training 2012/13. OPTA Sportsdata Madrid.

Pollard, R., Armatas, V., Hojjat, S., & Sani, Z. (2017). Home advantage in professional football in Iran - Differences between teams, levels of play and the effects of climate. *International Journal of Science Culture and Sport,* 5(4), 328-339. https://doi.org/10.14486/IntJSCS696

Rampinini, E., Impellizzeri, F. M., Castagna, C., Coutts, A. J., & Wisløff, U. (2009). Technical performance during soccer matches of the Italian Serie A league: Effect of fatigue and competitive level. *Journal of Science and Medicine in Sport*, *12*(1), 227–233. https://doi.org/10.1016/j.jsams.2007.10.002

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. https://doi.org/10.1080/02640410050120078

Russell, M., Rees, G., & Kingsley, M. I. C. (2013). Technical demands of soccer match play in the English championship. *Journal of Strength and Conditioning Research*, *27*(10), 2869–2873. https://doi.org/10.1519/JSC.0b013e318280cc13

Sainz de Baranda, P., Adán, L., García-Angulo, A., Gómez-López, M., Nikolic, B., & Ortega-Toro, E. (2019). Differences in the offensive and defensive actions of the goalkeepers at women's FIFA World Cup 2011. *Frontiers in Psychology*, *10*, 223. https://doi.org/10.3389/fpsyg.2019.00223

Sampaio, J., Janeira, M., Ibáñez, S., & Lorenzo, A. (2006). Discriminant analysis of game-related statistics between basketball guards, forwards and centres in three professional leagues. *European Journal of Sport Science*, *6*(3), 173–178. https://doi.org/10.1080/17461390600676200

Sarmento, H., Marcelino, R., Anguera, M. T., CampaniÇo, J., Matos, N., & LeitÃo, J. C. (2014). Match analysis in football: A systematic review. *Journal of Sports Sciences*, *32*(20), 1831–1843. https://doi.org/10.1080/02640414.2014.898852

Sarmento, H., Pereira, A., Matos, N., Campaniço, J., Anguera, T. M., Leitão, J., & Anguera, M. T. (2013). English Premier League, Spain's La Liga and Italy's Serié A - What's different? *International Journal of Performance Analysis in Sport*, *13*(3), 773–789. https://doi.org/10.1080/24748668.2013.11868688

Scott, P., & Tweedale, A. (2018). Premier League transfer business analysed: spending falls as shorter summer window comes into effect. Retrieved February 28, 2019, from https://www.telegraph.co.uk/football/2018/08/10/premier-league-transfer-business-analysed-spending-falls-shorter/

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.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Boston, MA: Pearson.

UEFA Club Licensing and Financial Fair Play Regulations. (2015). Retrieved from https://www.uefa.com/MultimediaFiles/Download/Tech/uefaorg/General/02/26/77/91/2267791_DOWNLOAD.pdf

Vaeyens, R., Lenoir, M., Williams, A. M., & Philippaerts, R. M. (2008). Talent identification and development programmes in sport. *Sports Medicine*, *38*(9), 703–714. https://doi.org/10.2165/00007256-200838090-00001

Vaeyens, R., Philippaerts, R. M., & Malina, R. M. (2005). The relative age effect in soccer: A match-related perspective. *Journal of Sports Sciences*, *23*(7), 747–756. https://doi.org/10.1080/02640410400022052

Wiemeyer, J. (2003). Who should play in which position in soccer? Empirical evidence and unconventional modelling. *International Journal of Performance Analysis in Sport*, *3*(1), 1–18. https://doi.org/10.1080/24748668.2003.11868269

Williams, A., & Reilly, T. (2000). Talent identification and development in soccer. *Journal of Sports Sciences*, *18*(9), 657–667. https://doi.org/10.1080/02640410050120041

Williams, R. (2015). Multicollinearity. University of Notre Dame. Retrieved from https://www3.nd.edu/~rwilliam/

Wilson, J. G. (2017). Talent identification and development of youth athletes in football. *Journal of Australian Strength & Conditioning*, *25*(7), 73–78.

Yi, Q., Jia, H., Liu, H., & Gómez, M. Á. (2018). Technical demands of different playing positions in the UEFA Champions League. *International Journal of Performance Analysis in Sport*, *18*(6), 926–937. https://doi.org/10.1080/24748668.2018.1528524

Appendix A

Due to the length of the code written to automate the process of data collection from the two websites, the entirety of the Python code has been uploaded to GitHub (https://github.com/X(anonymized for purpose of submission)) for easy reference.