The Relationship Between Pre-Season Functional Screening Test Performance and Coaches' Perception of Soccer Performance in Adult Male Players

Fahad Alfarraj [1] **, Jaquelin Bousie** [2] **, Jeremy Witchalls** [3] **, Phil Newman** [4]

1. Physiotherapy Department, Prince Sultan Military Medical City, Riyadh, SA ,

2. Faculty of Health, UC Research Institute of Sport and Exercise Science (UCRISE), University of Canberra, ACT 2601, AU ,

3. Faculty of Health, UC Research Institute of Sport and Exercise Science (UCRISE), University of Canberra, ACT 2601,, AU ,

4. Faculty of Health, UC Research Institute of Sport and Exercise Science (UCRISE), University of Canberra, ACT 2601, AU

Abstract: Coaches typically play a central role in team performance and player selection, but their assessments can also be valuable for the medical team in injury prevention and recovery. This research aimed to evaluate whether soccer coaches' evaluations of their players' physical abilities align with performance testing during pre-season. The players were rated subjectively by two coaches independently, focusing on various aspects such as technical, tactical, physical, and psychological skills. Ratings were given on a scale of 0 to 100, based on the coaches' perceptions of top players in those positions globally. The mean score of the coaches' ratings was used for each player. The study used the Intra-class Correlation Coefficient (ICC) to measure the reliability of inter-coach ratings, while players' scores on common functional tests were assessed independently by the medical staff. Decision tree analysis was conducted to determine the association between coaches' ratings and functional testing scores, as well as to identify cut-off values for differentiating higher and lower coach ratings. Sixty-three male professional soccer players from the Saudi Professional League participated voluntarily. The ICC values ranged from 0.73 to 0.79, indicating good to excellent agreement between coaches. The analysis showed that functional performance scores and coach ratings agreed in 86% of cases, with 88% precision and 91% recall. The algorithm correctly identified 88.4% of players rated high performers by coaches and 80% of lower-rated players. Cut-off scores were determined based on specific functional test results. For instance, players scoring above certain thresholds on tests like the Y-balance test and triple medial hop were more likely to receive higher coach ratings. The study revealed a solid alignment between coaches' subjective evaluations of physical abilities and objectively measured functional performance tests. These results hold the potential for aiding in player selection, establishing preparation standards, and facilitating players' return to play after injury.

Keywords: Functional tests, coach, performance, rating

-------------------------------------X-------------------------------------

INTRODUCTION

During the previous ten years, a concept that has been central in sporting research has been the focus on relationships that develop between a coach and athletes [1]. Indeed, it is commonly agreed that coaches are active in the development and advancement of athletes, as they can influence players (whether that be positively or negatively) and provide motivation through their own experiences in the sport [2]. Correspondingly, it has been shown that positive results within these sporting relationships are linked to quality outcomes [3].

In addition, it is currently known that sufficient coaching interaction helps athletes develop better long-term planning through a coach's expertise, which is central to their development in the sport [4]. Hence, the development and improvement of any athlete's performance need to stem from relevant and beneficial knowledge that is imparted from the coach [5]. The process under a coach, in its traditional format, includes team and player analysis, as well as planning and conducting specific interventions in regards to four major parts of player performance: physical, mental, tactical, and technical skills [6]. Moreover, the assessment of performance is directly relevant within a clinical setting, as it helps in injury diagnoses and prognosis, together with other medical conditions. It is also utilised to analyse how medical or exercise interventions can prove beneficial and therapeutic in returning to the relevant sport [7]. Coaches' ratings were placed into individual categories in the current study, which comprised of physical, technical, tactical, and psychological. Separately, students' physical capabilities can be rated reliably from physical education assessments, as has been demonstrated from a study on youth field hockey players [8]. Performance classification within any team sport can involve different measures of a variety of variables, which include: anthropometric testing; measurements of different physical attributes support the ability to play [9-13]; skill testing, and measures of how the sport is performed in regards to particular tasks in both training and match-play [14-16]; aspects of psychology [17, 18]; performance through team tactics; and the roles and/or positions within the whole team structure [19, 20]. Both general and sport-specific player capabilities can be evaluated through physical testing, although individual results are not always used in the prediction of match-play performance, due to an individual's competitive performance comprising a mixed complex nature [21].

Different studies support numerous performance measurements, including sports science measurement utilisation within the remit of both physical testing and in-match-play analysis, together with skills assessment of both technical and tactical measures [22]. Accordingly, it has been stated that a mixed testing approach is more valid in the evaluation of an individual's performance [23]. Besides that, machine learning (ML) is often useful in complex tasks that are undertaken within suitable timeframes [24]. Therefore, by using a ML environment, the study aimed to assess whether soccer coaches' assessments of their players' physical skills (in the domains: technical, tactical, physical, and psychological) are associated with the players' physical performance on formal performance testing during pre-season.

METHODS

The current study incorporated a coaching survey regarding how the coaches' subjective expert opinions perceive different movements and skills of players, with two different coaches performing the evaluation independently. Movement skill was rated using a 1-100 scale. General individual skill was assessed during training sessions and in match-play, where the coaches provided a mark out of 100 for the categories. The score of 100 was rated, comparing the coaches' perceptions of the world's leading players in those positions. The ratings were the mean of the two coaches' observations to produce one set rating for each participant.

Measures

The questionnaire included several sections in which the coaches used their expert subjective opinion to rate players' performances on various soccer performance factors. It was also determined that ten years of

required experience are necessary to determine experienced coaches from those declared novices [25]. Two qualified coaches ranked "experienced" rated players into four categories: physical, technical, tactical, and psychological; these were established using standardised definitions. Firstly, the physical measure assesses whether an athlete is physically ready for a game regarding their intrinsic fitness, strength, and neuromuscular control. Secondly, the technical item measures how a player is able to perform their specific game-related position and skills (both defence and offence) and their off-ball play. Thirdly, the tactical element measures the ability to perform general tactics, including ball possession and when not in possession. Fourthly, the psychological stage indicates an individual's capacity during the game and a level of mental strength, confidence, and emotional commitment when playing in that position.

Facilities and Participant's Preparation

The study assessed football players from the Saudi Professional League, utilising their respective team facilities without additional funding for conducting the tests. Hence, funding was not required to undertake these tests. Additionally, as there was a requirement to collate identified and confidential information, the analysis in the study and the results were accessed purely by the research team. Further, in accordance with a risk assessment and mitigation plan, the medical team used standard COVID-19 screening for each participating player before coming to the test lab. Each test session also comprised standard operating procedures to standardise the tests' administration. Each participating athlete was asked to ride on an exercise bicycle with minimal resistance for 5 minutes prior to starting the test, to function as the warm-up and help to prevent any potential injuries during the analysis [26]. Subsequently and in a random order, the single-leg functional tests included Y Balance, Triple Medial Hop, Triple Forward Hop, and Hexagon Agility tests. While conducting the tests, the subjects wore training shoes and performed on a surface made of rubber tiles. A 2-5 minutes recovery time was given to the participants between each test and the following one. Three practice trials, which were averaged for data analysis, were undertaken to acclimate the athletes to the tests and reduce the learning effect that may occur [27]. The participants' Y Balance reach distances were recorded in centimetres (cm) and then normalised by dividing by each individual's leg length (anterior superior iliac spine to the medial malleolus' distal tip) [28].

Tests

Y Balance Test – the players were asked to stand on one leg in a central grid, with their big toe positioned at the starting line. Each participant was asked to maintain the single-leg stance, and the player simultaneously needed to reach with his free leg in the anterior, posteromedial and posterolateral directions. Each direction was completed in separate trials. A tape measure was used to determine the maximal reach distance, which was completed by marking the farthest reached point of the distal part of the foot [29].

The Triple Medial Hop Test – The players were asked to place one of their feet perpendicular to the measuring tape's endpoint. From a starting position standing on one leg, the player hopped three times as far as possible in a medial direction. The measurement in centimetres was taken of the distance between the heel's lateral surface at the starting and final positions [30]. In comparison, the Triple Forward Hop Test [31] required three consecutive straight-line maximal forward hops, with a measurement taken in

centimetres of the distance from the toe of the original take-off point to the final position.

The Hexagon Hop Test [32] – The players stood on their test leg in a hexagon with 60cm sides marked on the gym floor (including a 40cm circle marked in the centre) in order to start the test, and then hop out and back into the different sides of the hexagon in a sequence (moving clockwise). The participants faced forwards throughout the test. The observer counted the number of out-and-back jumps that the participants completed without touching the lines on the hexagon and ensured that they had hopped back sufficiently into the hexagon to make contact with the circle lines. The test lasted for 10 seconds for each direction, with a short rest before being repeated in the anticlockwise direction. A combination of the number of accurate hops completed in each direction determined the total score for each leg. A record showing the performance on individual legs enabled the analysis of performance asymmetry between the legs on all tests, with escalating power and fatigue throughout the testing sequence.

DATA ANALYSIS

The Intra-class Correlation Coefficient (ICC) was utilised in SPSS (version 27.0) to assess the reliability of the inter-coach ratings. A two-way random-effects' model were used with the measure of consistency setting. ICCs were defined as <0.40 (poor), 0.40–0.75 (fair to good), and >0.75 (excellent) [33]. Datasets were imported to Orange Data Mining (Version 3.33) [34] (see **Figure 1**). ML metrics, such as Area Under the Curve (AUC), Classification Accuracy (CA), and precision and recall scores, were utilised to summarise the model's performance. Separately, decision tree analysis was deployed by using the depth of 6 levels to determine: 1) How closely coaches' ratings of physical aptitude are associated with functional testing scores; and 2) What cut-off values best discriminated between higher and lower coach ratings based on mean scores?
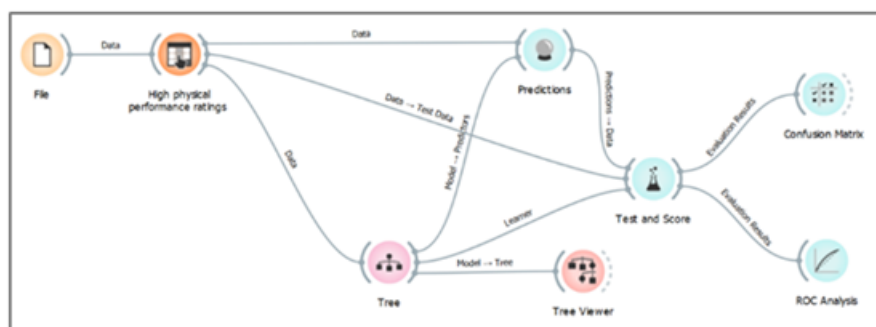

Figure 1. Analysis process for the high physical performance ratings

RESULTS

Descriptive data were recorded (means, SDs) for both player characteristics and the pre-season functional screening tests (PSFSTs) (see **Table 1**).

Table 2 includes ICCs with 95% CI for inter-coach ratings reliability levels for the all-sports' performance factors. The ICC values ranged from 0.73 – 0.79 for the sports performance factors, which indicated levels of good to excellent agreement between coaches. Meanwhile, the tree model (see) demonstrated that functional performance scores could be used to distinguish high versus low-rated players with 86%

accuracy, precision 88% and recall 91%. Error! No bookmark name given.) shows that the algorithm using functional testing scores rated 20% of players as less physically capable when their coaches rated them as high performers. The decision tree correctly rated 88.4% of players classified as high physical performers by their coaches, and 80% of lower-rated players. The decision trees (see) provided cut-off scores, where high physical performance ratings from the coaches were given to 42 out of 63 players. The cut-off scores that best discriminated between higher and lower coach ratings were: average bilateral anterior normalised Y-balance test greater than 63.7 norm-cm, average bilateral triple medial hop between 408.3 cm and 481.7 cm; and average bilateral posterolateral normalised Y-balance test greater than 88.2 norm-cm.

Table 1. Descriptive Statistics

	N †	Mean	Std. Deviation
Player characteristics			
Age (years)		23.08	1.34
Height (cm)	63	176.16	6.56
Weight (kg)		72.90	7.47
BMI (kg/m²)		23.47	1.83
Tests ‡			
Anterior normalised YBT (norm-cm)		69.50	6.66
Posteromedial normalised YBT (norm-cm)		94.79	7.99
Posterolateral normalised YBT (norm-cm)	63	90.20	8.26
Triple medial hop (cm)		451.21	45.54
Triple forward hop (cm)		527.01	52.03
Hexagon agility (rep per 10 sec)		13.32	1.95
† Population size			
‡ Average measures of both legs			

Table 2. Intraclass Correlation Coefficient

Inter-coach ratings, Performance factor	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Physical	0.76	0.63	0.85	7.29	62	62	0.000
Technical	0.75	0.61	0.84	6.85	62	62	0.000
Tactical	0.79	0.68	0.87	8.65	62	62	0.000
Psychological	0.73	0.59	0.83	6.44	62	62	0.000

Table 3. Precision, recall, and confusion matrix of Tree model the physical performance part of the

coach ratings

Tree Model	
Precision	Recall
0.884	0.905

Confusion Matrix

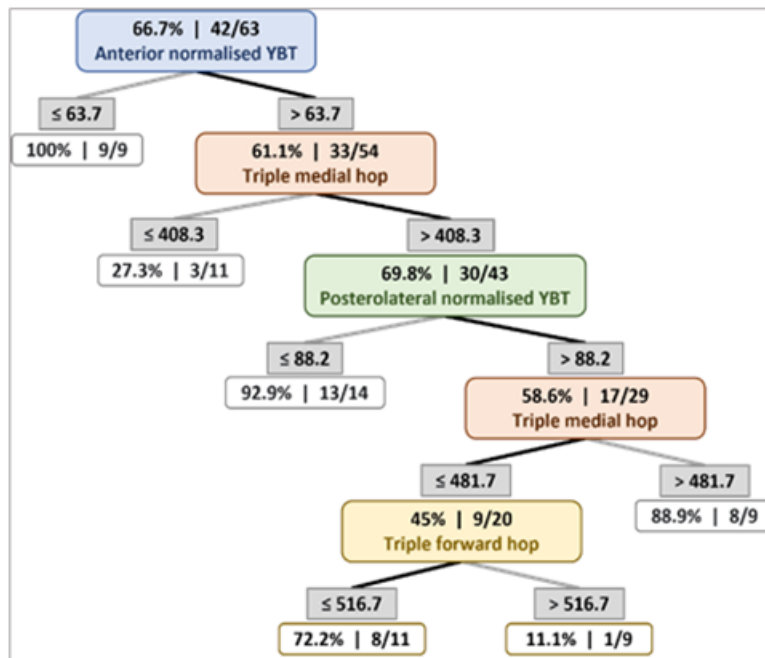|		Predicted		
		Low performer	High performer	
Actual	Low performer	80%	11.6%	21
	High performer	20%	88.4%	42
		20	43	63

DISCUSSION

Figure 2. Discriminated pre-season functional screening tests scores between higher and lower coach ratings

The current study investigates the relationship between PSFSTs and the performance of soccer players, which involved a survey of two coaches and how their subjective opinions as experts regarding athletes'

sports performance (technical, tactical, physical and psychological) contribute to this dynamic. The findings indicate a level of good to excellent inter-coach rating reliability. The study also shows good agreement between coach ratings of performance and a ML model including PSFSTs.

The results indicate functional performance scores could be used to distinguish high versus low-rated players, which could provide the various forms of effort that soccer requires that can be shown through the field tests conducted on strength, balance, and endurance. Accordingly, the higher-level performing individuals have increased leg control and were able to perform at a better level of skill.

In the modern process of analysis, ML has been able to improve knowledge levels, with computers taking the role of humans and feeding data over time autonomously [35]. ML in the field of soccer has been undertaken using a variety of predictive algorithms, with the majority using 'decision trees' [36]. Hence, decision trees were selected as base classifiers, as they are able to create understandable models that can provide decision thresholds or cut-off scores. This generates a model of performance characteristics that sport practitioners can implement in into programmes. As far as is known, this is the only research study which has focused on soccer and examined the relationship between the opinion of coaches on player performance and measures of players' performance. Thus, it may be interesting to compare results with previous studies. A recent study [37] highlighted that screening tests which have been selected from the standard musculoskeletal tests and have been implemented in field hockey squads correlate with coaches' ratings of top and lower game performers. Conversely, it has been suggested in field hockey research that top-level players present higher levels of technical and tactical variables, although these were not shown in screening or physical tests [15]. A different study [38] demonstrated that strength and muscle power correlated with more talented players amongst other variables; whilst flexibility and particular anthropometric measures related to higher-performing players [39].

The YBT appeared in the current study as a vital component of the decision tree model with reach directions of both anterior and posterolateral. This result partially agreed with a previous study [37], which found that posteromedial and posterolateral YBT was associated with physical and technical ratings in top male hockey performers. There was no relationship found with coaches' ratings in triple forward hop or in the hexagon agility test in that study. Nevertheless, value may be present in relation to the management of injuries or predictions.

A coach's experience, together with in-depth knowledge of specific sporting requirements help to determine players' performance levels and quality, as it is possible to observe the level that a player can achieve. This can be viewed from a combination of performance data and the identification of performance factors. The current research has demonstrated that players' physical characteristics are correlated to their own perceptions and the views of their coaches, which makes it possible to implement particular targeted training strategies that focus on players' physical performance.

Despite the findings from the current study, it should be noted that the number of participants is small. Accordingly, it has been stated [40] that when a research sample size is sufficiently large, it is possible to divide the data into different sets of training and validation. From these training datasets, it subsequently becomes possible to develop a decision tree model, together with a validation dataset that enables a relevant required tree size that will achieve the ideal final model. Factors, including anthropometric

measurements, were not taken into account, as the cohort had been pre-selected to a level that acknowledged their level of talent. Performance measures that were required in this respect had to be relevant to the identification of the differences between top-performing individuals and lower-level performance comparisons in a universally elite cohort, with ratings providing more than attainment measures.

CONCLUSION

Coach rating scales are an efficient measure of qualitative information and are able to incorporate context and sport-specific aspects. This is the initial step in supporting their continued use in team sports. Further, findings from the decision tree demonstrate that physical performance does appear to be related to coaches' scores of players. This could also be utilised to help in the selection of players, preparation criteria, and players returning to play following injury, as well as providing a base for further research to develop the predictive ability of the test battery.

References

1. Jowett, S. and P. Wylleman, Interpersonal relationships in sport and exercise settings: Crossing the chasm. Psychology of Sport & Exercise, 2006. 2(7): p. 119-123.

2. Bruner, M.W., J. Hall, and J. Côté, Influence of sport type and interdependence on the developmental experiences of youth male athletes. European Journal of Sport Science, 2011. 11(2): p. 131-142.

3. Rhind, D.J. and S. Jowett, Initial evidence for the criterion-related and structural validity of the long versions of the Coach–Athlete Relationship Questionnaire. European Journal of Sport Science, 2010. 10(6): p. 359-370.

4. Leite, N., J. Baker, and J. Sampaio, Paths to expertise in Portuguese national team athletes. Journal of Sports Science & Medicine, 2009. 8(4): p. 560-566.

5. Côté, J. and W. Gilbert, An integrative definition of coaching effectiveness and expertise. International Journal of Sports Science & Coaching, 2009. 4(3): p. 307-323.

6. Carling, C., T. Reilly, and A.M. Williams, Performance assessment for field sports. 2008: Routledge.

7. Winter, E., et al., Rationale. Sport and Exercise Physiology Testing Guidelines, 2007. 1: p. 7-10.

8. Elferink-Gemser, M., et al., Relation between multidimensional performance characteristics and level of performance in talented youth field hockey players. Journal of Sports Sciences, 2004. 22(11-12): p. 1053-1063.

9. Ré, A.H.N., U.C. Corrêa, and M.T.S. Böhme, Anthropometric characteristics and motor skills in talent selection and development in indoor soccer. Perceptual and Motor Skills, 2010. 110(3): p. 916-930.

10. Bishop, D.J. and O. Girard, Determinants of team-sport performance: implications for altitude training by team-sport athletes. British Journal of Sports Medicine, 2013. 47(Suppl 1): p. i17-i21.

11. Cochrane, D. and S. Stannard, Acute whole body vibration training increases vertical jump and flexibility performance in elite female field hockey players. British Journal of Sports Medicine, 2005. 39(11): p. 860-865.

12. Cressey, E.M., et al., The effects of ten weeks of lower-body unstable surface training on markers of athletic performance. The Journal of Strength & Conditioning Research, 2007. 21(2): p. 561-567.

13. Hrysomallis, C., Balance ability and athletic performance. Sports Medicine, 2011. 41(3): p. 221-232.

14. Gabbett, T.J. and B. Georgieff, The development of a standardized skill assessment for junior volleyball players. International Journal of Sports Physiology and Performance, 2006. 1(2): p. 95-107.

15. Elferink-Gemser, M.T., et al., Multidimensional performance characteristics and standard of performance in talented youth field hockey players: A longitudinal study. Journal of Sports Sciences, 2007. 25(4): p. 481-489.

16. Sunderland, C., et al., The reliability and validity of a field hockey skill test. International Journal of Sports Medicine, 2006. 27(05): p. 395-400.

17. Landy, F.J. and J.L. Farr, Performance rating. Psychological Bulletin, 1980. 87(1): p. 72.

18. Calder, J.M. and I.N. Durbach, Decision support for evaluating player performance in rugby union. International Journal of Sports Science & Coaching, 2015. 10(1): p. 21-37.

19. Sullivan, C., et al., Factors affecting match performance in professional Australian football. International Journal of Sports Physiology and Performance, 2014. 9(3): p. 561-566.

20. Tromp, E.Y., et al., "Let's Pick Him!": Ratings of Skill Level on the Basis of in-Game Playing Behaviour in Bantam League Junior ICE Hockey. International Journal of Sports Science & Coaching, 2013. 8(4): p. 641-660.

21. Svensson, M. and B. Drust, Testing soccer players. Journal of Sports Sciences, 2005. 23(6): p. 601-618.

22. Phillips, E., et al., Expert performance in sport and the dynamics of talent development. Sports Medicine, 2010. 40(4): p. 271-283.

23. Lames, M. and T. McGarry, On the search for reliable performance indicators in game sports. International Journal of Performance Analysis in Sport, 2007. 7(1): p. 62-79.

24. Pang, B., L. Lee, and S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002.

25. Abraham, A., D. Collins, and R. Martindale, The coaching schematic: Validation through expert coach consensus. Journal of Sports Sciences, 2006. 24(06): p. 549-564.

26. Woods, K., P. Bishop, and E. Jones, Warm-up and stretching in the prevention of muscular injury. Sports Medicine, 2007. 37(12): p. 1089-1099.

27. Robinson, R.H. and P.A. Gribble, Support for a reduction in the number of trials needed for the star excursion balance test. Archives of Physical Medicine and Rehabilitation, 2008. 89(2): p. 364-370.

28. Gribble, P.A. and J. Hertel, Considerations for normalizing measures of the Star Excursion Balance Test. Measurement in Physical Education and Exercise Science, 2003. 7(2): p. 89-100.

29. Gribble, P.A., J. Hertel, and P. Plisky, Using the Star Excursion Balance Test to assess dynamic postural-control deficits

30. Kivlan, B.R., et al., Reliability and validity of functional performance tests in dancers with hip dysfunction. International Journal of Sports Physical Therapy, 2013. 8(4): p. 360-369.

31. Noyes, F.R., S.D. Barber, and R.E. Mangine, Abnormal lower limb symmetry determined by function hop tests after anterior cruciate ligament rupture. The American Journal of Sports Medicine, 1991. 19(5): p. 513-518.

32. Witchalls, J.B., et al., Functional performance deficits associated with ligamentous instability at the ankle. Journal of Science and Medicine in Sport, 2013. 16(2): p. 89-93.

33. Fleiss, J., The design and analysis of clinical experiments. New York: Willey and Sons, 1986: p. 1-15.

34. Demšar, J., et al., Orange: data mining toolbox in Python. The Journal of Machine Learning Research, 2013. 14(1): p. 2349-2353.

35. Nasteski, V., An overview of the supervised machine learning methods. Horizons. b, 2017. 4: p. 51-62.

36. Rico-González, M., et al., Machine learning application in soccer: A systematic review. Biology of Sport, 2022. 40(1): p. 249-263.

37. Stokes, M., et al., Are Intrinsic Factors useful in predicting Risk of Injury and Performance in Field Hockey, in Research Institute for Sport and Exercise. 2020, University of Canberra.

38. Keogh, J.W., C.L. Weber, and C.T. Dalton, Evaluation of anthropometric, physiological, and skill-related tests for talent identification in female field hockey. Canadian Journal of Applied Physiology, 2003. 28(3): p. 397-409.

39. Nieuwenhuis, C.F., E.J. Spamer, and J.H.v. Rossum, Prediction function for identifying talent in 14-to 15-year-old female field hockey players. High Ability Studies, 2002. 13(1): p. 21-33.

40. Song, Y.-Y. and L. Ying, Decision tree methods: applications for classification and prediction. Shanghai Archives of Psychiatry, 2015. 27(2): p. 130