



VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

Validation of the Australian Football League Player Ratings

This is the Accepted version of the following publication

McIntosh, Sam, Kovalchik, Stephanie and Robertson, Samuel (2018)
Validation of the Australian Football League Player Ratings. *International journal of Sports Science and Coaching*. ISSN 1747-9541

The publisher's official version can be found at
<http://journals.sagepub.com/doi/full/10.1177/1747954118758000>
Note that access to this version may require subscription.

Downloaded from VU Research Repository <https://vuir.vu.edu.au/37076/>

Original Article

Corresponding Author:

Sam McIntosh, Institute of Sport, Exercise & Active Living, Victoria University, Melbourne, Victoria, Australia.

Email: sam.mcintosh@live.vu.edu.au

Validation of the Australian Football League Player Ratings

Sam McIntosh^{1,2}, Stephanie Kovalchik¹ & Sam Robertson^{1,2}

¹ Institute of Sport, Exercise & Active Living, College of Sport and Exercise Sciences, Victoria University, Melbourne, Australia

² Western Bulldogs, Melbourne, Australia

Abstract

This study investigated the validity of the official Australian Football League (AFL) Player Ratings system. It also aimed to determine the extent to which the distribution of points across the 13 rating subcategories could explain AFL match outcome. Ratings were obtained for each player from AFL matches played during the 2013-2016 seasons, along with the corresponding match outcome (Win/Loss and score margin). The values for each of the 13 subcategories that comprise the ratings were also obtained for the 2016 season. Total team rating scores were derived as an objective team outcome for each match. Percentage agreement and Pearson correlational analyses revealed that winning teams displayed a higher total team rating in 94.2% of matches, and an association of $r = 0.96$ (95% confidence interval of 0.95-0.96) between match score margin and total team rating differential, respectively. A PART analysis resulted in seven rules capable of determining the extent to which relative contributions of rating subcategories explain Win/Loss at an accuracy of 79.3%. These models support the validity of the AFL Player Ratings system, and its use as a pertinent system for objective player analyses in the AFL.

Keywords

performance analysis; decision support; operations research; player evaluation; team sport.

Introduction

Performance analysis is used within sporting organisations to support decision-making processes relating to an individual or team's performance.^{1, 2} In many professional sports, various rating systems have been proposed with the aim of encapsulating player or team performance on a quantitative scale.³⁻⁶ In individual sports, the match result can be used as an objective outcome to directly compare against performance.⁵ Similarly in team sports such as baseball, individual performance has been objectively quantified as a result of direct player actions.^{7, 8} However, rating individuals within invasion team sports such as Australian Rules football (AF) and football is more complex.⁹ This is in part due to the absence of objectively quantifiable outcomes that emanate directly from player actions, but also the dynamic nature, varied individual roles and complex interactions which exist between individuals in these sports.^{2, 10}

Within the elite competitions of many invasion team-based sports, an increase in the collection and reporting of performance data has led to the existence of more detailed and comprehensive performance rating systems.¹¹ This has in turn resulted in those responsible for making organisational decisions become more reliant on performance data to make inferences about player performance and support their decision making processes.¹¹ McHale et al.⁵ developed a player performance index which is used within the top two tiers of English football. This system rates the performance of individual players on a quantitative scale, based on their contributions to weighted sub-indices. Similarly in basketball, the player efficiency rating is a broadly used objective rating system which measures a players temporally-adjusted productivity based on positive and negative actions, and their outcomes.⁶

Australian Rules football is an invasion team sport played on an oval field between two opposing teams consisting of 22 players each (18 on the field and four interchange). The ball is moved about the field by kicking, handballing or running with the ball, with scoring achieved by kicking the ball between large goal posts located at either end of the field. Within the elite competition of AF, the Australian Football League (AFL), various subjective rating systems have been proposed that quantify an individual's match performance. However, these are susceptible to biases, such as personal views and emotional reflection, which are known to accompany such subjective analyses.^{12, 13} For instance, the AFL Coaches Association awards a champion player each year. Votes for this are cast following each match by the senior coaches from both competing teams on the most influential players from their respective match.

From an objective perspective, Heasman et al.¹⁴ created a player impact rating by attributing numerical values to performance actions relative to their perceived worth, weighting these values according to match situation and then adjusting relative to a players time on ground. Following the release of the novel *Moneyball*,¹⁵ Stewart et al.¹⁶ determined whether similar statistical methods could be applied to the AFL. Using data from five seasons, they created an 11-variable player ranking model by identifying the most important performance actions and then including those with the strongest statistical relationship to team winning margin. The 'AFL Player Rankings', which is produced by statistics provider Champion Data Pty Ltd and is the system used by the fantasy competition *SuperCoach* (www.supercoach.heraldsun.com.au), takes a similar approach to that of Stewart et al.¹⁶ however extends their model to include over 100 variables.¹⁷ To date there has been no external research to evaluate the validity of these systems.

Recently, a new alternative to the abovementioned systems has been proposed; the ‘AFL Player Ratings’ (<http://www.afl.com.au/stats/player-ratings/ratings-hub>). Produced by Champion Data, it is an objective system based on the principle of field equity, where a player’s actions are quantified relative to how much their actions increase or decrease their team’s expected value of the next score.¹⁸ For example, when a player obtains the ball in a contested situation a long distance away from their attacking goal, the expected value of next score is likely to be low (or negative, meaning in the given situation the opposition is more likely to score). Conversely, if a player receives the ball uncontested, with minimal pressure and is close to their own goal, the expected value of the next score will be high. This expected value is based on contextual information relating to each possession (i.e., pressure from opponents, field position, time of the match), and is determined by the outcomes from every possession collected from all AFL matches preceding back to the 2004 season.¹⁸ Furthermore, the rating points awarded to (or taken from) a player for each action falls into to one or more categories which describe the nature of the action. These categories are defined in Table 1.

The primary aim of this study was to determine the construct validity of the AFL Player Ratings system, using data collected from the 2013-2016 AFL seasons. The secondary aim was to determine the extent to which the distribution of points recorded by teams across the 13 rating subcategories could be used to explain AFL match outcome. This study incorporated two phases; the first phase focuses on the derived total team ratings, whilst the second phase considers the 13 player rating subcategories.

Table 1. Definitions of the 13 AFL Player Ratings subcategories used in this study.

Category	Subcategory	Description
Ball Winning	Stoppage	Points from possessions won pre-clearance at stoppages.
	Mid Chain	Points from possessions excluding those won at stoppages or as intercepts.
	Intercepts	Points from intercept possessions.
Ball Use	Run and Handball	Points from handballs, and ball carrying between the possession and handball.
	Field Kicks	Points from field kicks.
	Shots at Goal	Points from shots at goal.
	Kick-ins	Points from kick-ins.
Hitouts	Hitouts	Points from hitouts to advantage and points lost from hitouts to opposition. Neutral hitouts gain zero points.
Defence	Spoils	Points from spoils.
	Pressure	Points from pressure - including tackles and smothers.
Negatives	Frees Against	Points lost from frees against.
	50 m Penalties	Points lost from 50 m penalties against.
	Debits	Points lost from dropped marks, no pressure errors and missed tackles.

Methods

Phase One: Construct validity of the AFL Player Ratings system

Individual ratings data were obtained from Champion Data Pty Ltd, for all 827 matches played throughout the 2013-2016 AFL seasons. This included 22 matches from each team during the regular season rounds, as well as nine matches played throughout the finals series each season. One match was abandoned prior to play during the 2015 season. Match result was obtained for each match and expressed as a) *outcome* (Win/Loss) and b) *margin* (points score differential). Prior to data collection, the study was approved by the relevant human research ethics committee.

Total team ratings were derived for each match by accumulating the 22 individual player ratings from the same match. The total team rating was derived with the aim of providing an objective independent variable to be modelled against outcome and margin. This was completed for each of the AFL teams ($n = 18$), for each match played throughout the four seasons. Prior to statistical analysis, the four drawn matches that occurred throughout the 2013-2016 seasons were removed from the analyses.

For the remaining 823 matches, a percentage agreement analysis was used to construct a model explaining outcome as a function of higher total team ratings. Descriptive statistics (mean \pm standard deviation) of the total team match ratings were also collected across the four seasons to gauge the consistency of the system across seasons. In order to gauge the strength of total team rating differential as a continuous variable, a Pearson's correlation analysis was employed to determine the extent of its relationship with margin. This analysis was undertaken using the *Hmisc* package¹⁹ in the R statistical computing software version 3.3.2.²⁰ Correlations were obtained considering the entire dataset, as well as separately within team and across the whole competition for individual seasons, allowing for assessment of both inter-team and inter-season variations, respectively.

Phase Two: Relationships between the distribution of AFL Player Ratings subcategories and match result

To address the secondary aim, data from each subcategory of each individual's player ratings were obtained from Champion Data Pty Ltd. These analyses were limited to the 207 matches played throughout the 2016 AFL season due to data availability.

Descriptive statistics (mean \pm standard deviation) for all 13 subcategories were obtained across the season. In order to determine the relationship of each subcategory with match result, the total team ratings (as calculated in phase one) were broken down into separate contributions from each subcategory for each match. In order to allow for repeat observations across all teams and each round throughout the season, the data was then descriptively converted from its absolute format into a relative format.²¹ For example, if a team's match rating was 250 points, of which 30 points were attributed by the subcategory field kicks, then the team's relative contribution of field kicks for this particular match would be analysed as 12%.

To determine the extent to which the separate contributions from each subcategory related to outcome, a rule induction analysis was undertaken using the *RWeka* package²². A PART algorithm²³ was used to generate a list of rules capable of explaining outcome. For this analysis, overall classification accuracy (%) and 10-fold cross-validation accuracy were used as the two model performance measures. A number of parameters were trialled in the model development, with best performance based on the abovementioned measures obtained using a minimum of 20 instances in order for a node to split and minimum confidence set to 0.5.

Results

Phase One: Construct validity of the AFL Player Ratings system

Results from the percentage agreement analysis found winning teams had a higher total team rating in 94.2% of matches (775 of 823 instances), with winning teams averaging 232.1 ± 27.2 rating points across the four seasons and losing teams averaging 192.1 ± 25.5 . The density of total team ratings difference for winning teams is outlined in Figure 1, and the distribution of total team ratings for winning and losing teams across the four seasons is shown in Figure 2.

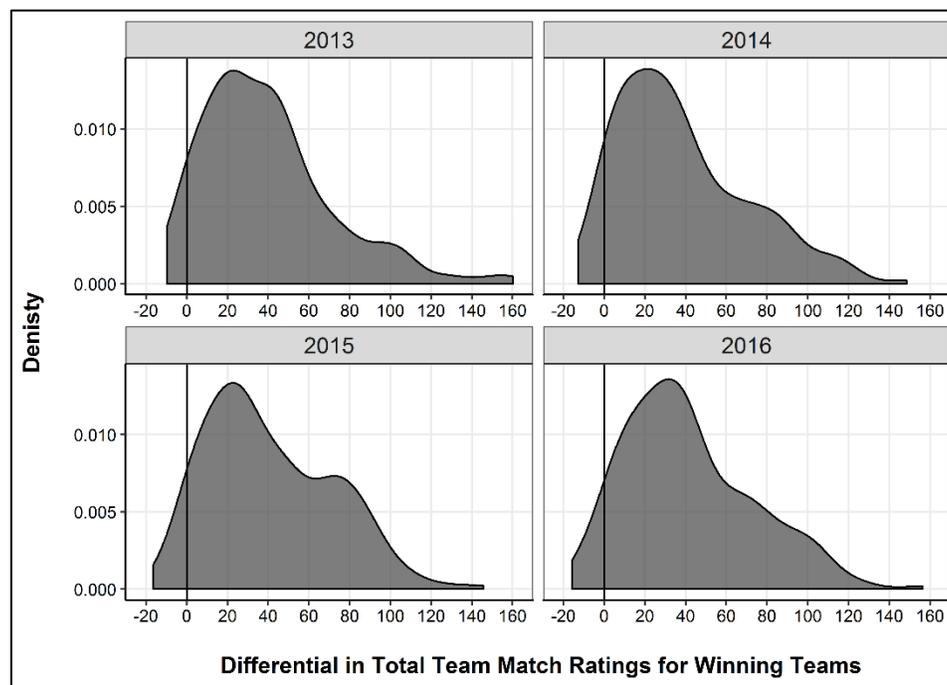


Figure 1. Density plot displaying the distribution of differentials in total team ratings for winning teams across the 2013-2016 seasons.

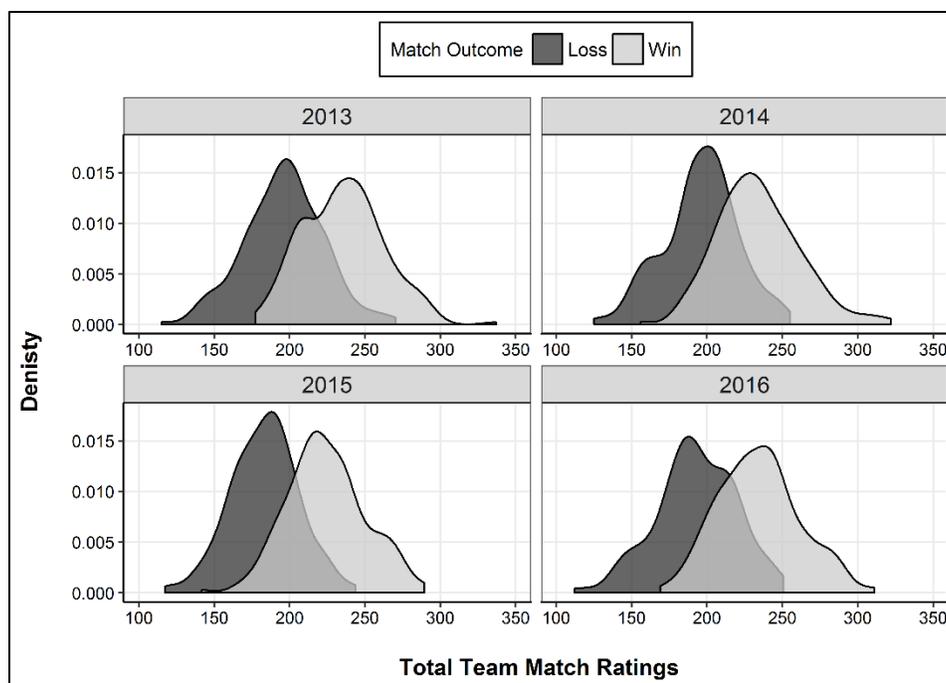


Figure 2. Density plot displaying the distribution of total team ratings across the 2013-2016 seasons.

The Pearson's correlation analysis indicated a strong association ($r = 0.96$, 95% confidence interval (CI) = 0.95-0.96) across the 18 teams in the competition, between margin and total team rating. The association for each of the 18 teams varied from $r = 0.92$ to $r = 0.97$ (95% CI = 0.89-0.96 and 0.95-0.98, respectively), and across the four seasons from $r = 0.95$ to $r = 0.96$ (95% CI = 0.94-0.95 and 0.96-0.97, respectively). The scatterplot shown in Figure 3 displays the linear association between margin and total team rating, indicating a homoscedastic distribution.

Phase Two: Relationships between the distribution of AFL Player Ratings subcategories and match result

Descriptive statistics relating to each of the subcategories and how their contributions differentiate between wins and losses are outlined in Figure 4. Results show that on average, winning teams had a higher contribution of team rating points in only the four subcategories which relate to ball use (run and handball, field kicks, shots at goal and kick-ins).

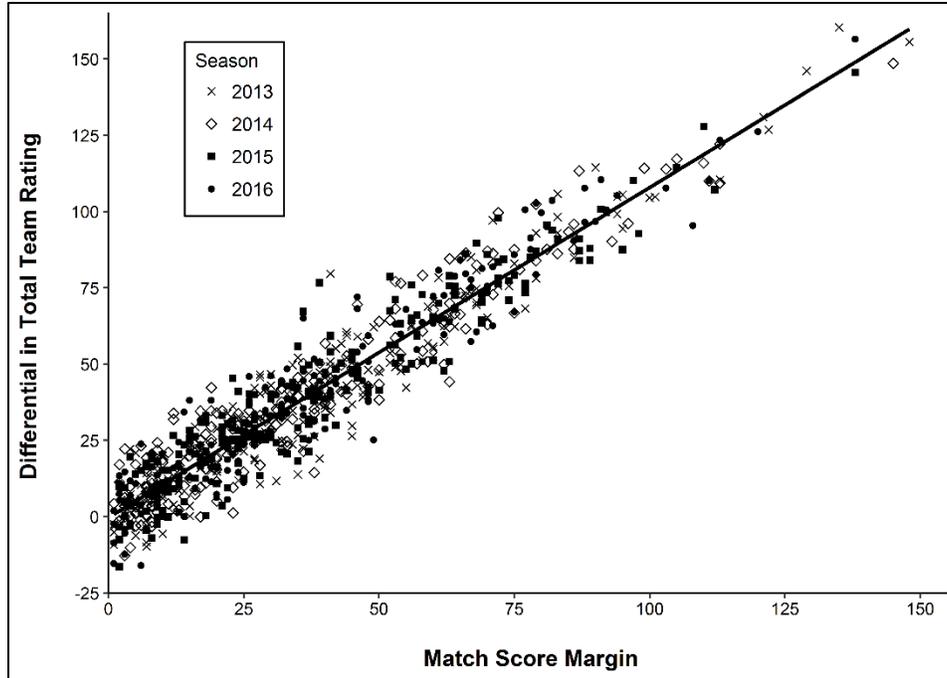


Figure 3. Scatterplot displaying the homoscedasticity of the distribution between margin and total team rating differentials across the 2013-2016 seasons.

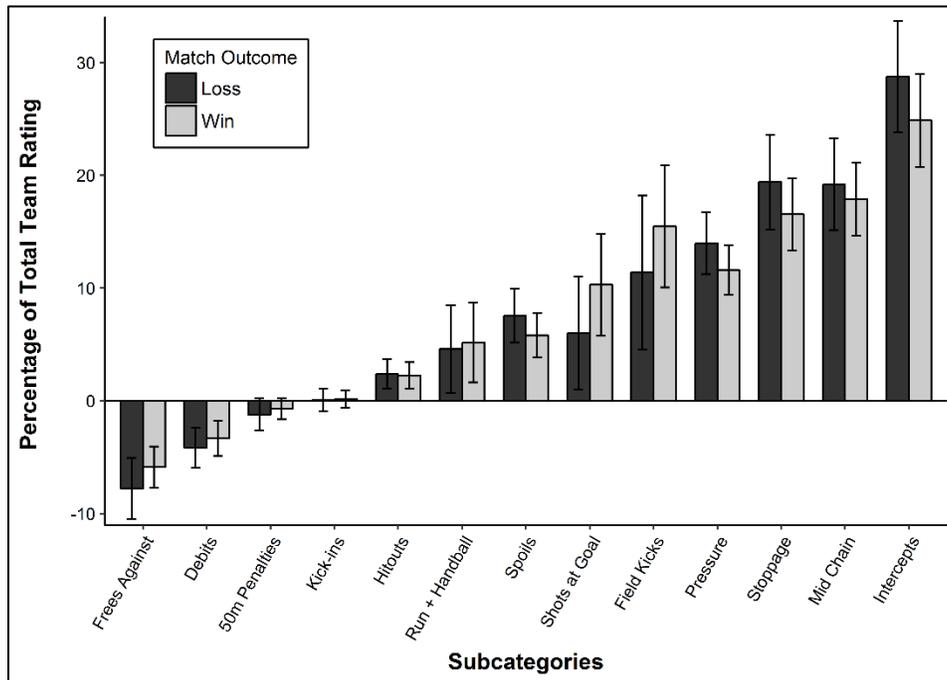


Figure 4. Histogram displaying the descriptive statistics (mean \pm standard deviation) of the relative contribution to total team rating from each of the subcategories across the 2013-2016 seasons.

The final PART model revealed seven rules explaining outcome at an accuracy of 79.3% (314 of 396 matches). The class accuracy rates in the data for each outcome were 70.2% for wins and 88.4% for losses. Results from the cross-validation revealed a decrease in classification accuracy of 6.1%, with an overall classification of 73.2% retained. The seven rules included in the model are presented in Table 2. The percentage values represent the contribution from each subcategory to team total rating.

Table 2. PART model explaining outcome from the relative contributions of each subcategory to team total rating.

Rule	Outcome	Correctly classified	Incorrectly classified
Rule 1 Intercepts > 27.1% AND Frees Against > -9.6% AND Shots at Goal > 4.2% AND Pressure > 11.6%	Loss	60	19
Rule 2 Shots at Goal ≤ 4.3%	Loss	87	14
Rule 3 Frees Against > -9.4% AND Spoils ≤ 4.9% AND Shots at Goal > 10.8%	Win	62	10
Rule 4 Spoils ≤ 4.9%	Win	62	4
Rule 5 Field Kicks < 15.2%	Loss	39	17
Rule 6 Pressure > 12.9%	Loss	48	9
Rule 7 Else	Win	38	9

Discussion

The primary aim of this study was to determine the construct validity of the AFL Player Ratings system. Phase one focused specifically on the ability of the AFL Player Ratings system to relate to match result when expressed in both a binomial (outcome) and continuous manner (margin). The findings revealed that the AFL Player Ratings system is strongly associated with match result irrespective of how it is expressed, suggesting that the system has good validity for assessing combined player performance in AF. The findings of the correlational analysis support the findings of the percentage agreement, highlighting that in the very low proportion of matches where agreement was not reached, both the margin and team total rating differential were both very small. The strength of these associations emphasise how incorporating considerations about the equity of a player's actions is a viable method of quantifying aggregated player performance.

Phase two focused on determining the extent to which the distribution of points across the 13 rating subcategories could be used to explain outcome. Descriptive statistics revealed that only those subcategories relating to ball use had a higher average contribution to team rating points by winning sides. This is likely a result of the ball use subcategories being the only four subcategories in which rating points can be both awarded and deducted. Therefore, contributions of points within these subcategories are further impacted by whether actions increase or decrease their team's expected value of the next score. Of the 13 subcategories included in the analysis, six are outlined in the PART model. Specifically, the model indicates a positive relationship between larger contributions of shots at goal and field kicks with successful outcome. This is unsurprising due to the function of scoring on match result, and the known relationship between maintaining ball possession and match result in AF,²⁴ respectively. Additionally, the positive relationship seen in

these two subcategories is again likely associated with the ability to both gain and lose rating points in these subcategories. Conversely, the model indicates an inverse relationship between larger contributions of pressure, spoils and intercepts with match outcome. Although points are awarded to players for actions in these subcategories, having above-average relative contributions in these subcategories reflects lower contributions in other subcategories, specifically those relating to ball use.

The absence of the remaining seven subcategories from the model is likely to be multifaceted. Specifically, for run and handball, kick-ins, hitouts, 50 m penalties and debits, a comparatively low overall contribution to team total ratings as well as small variation in mean values between wins and losses may have contributed to their absence. For stoppages and mid chain, despite a relatively higher overall contribution to team total ratings, their absence is potentially due to small variations to mean values between wins and losses.

As this study takes a specific focus on objective performance, an assumption was made that the sum of a team's parts (individual contributions) combine to create the result, therefore utilising successful team performance as an objective dependent variable. As such, this study focused on how the AFL Player Ratings reflect team results to provide a validation of the metrics construct. Heasman et al.¹⁴ took a similar approach in the validation of their player impact model, finding their team impact scores were higher in winning teams in 86.4% of matches (19 of 22 instances), and had a strong correlation with margin ($r = 0.85$). In comparison, the findings of both the percentage agreement and Pearson's correlation models in this study had stronger relationships with respect to match outcome and margin, respectively. A larger sample size was also used. Stewart et al.¹⁶ also considered score margin to identify which player statistics are most important in terms of their contribution to match outcome. Their findings indicate that kicks travelling more than 40 m and kicks that go directly to an opposition player have large positive and negative coefficients, respectively. Thus reiterating the findings of phase two in this study, indicating that actions relating to ball use have the largest impact on match outcome. It is not known as to whether the AFL Player Ratings displays higher construct validity comparative to popular fantasy football metrics, however future research may look to determine this. Though adopting a team approach for this validation was necessary, future research should look to assess the contribution of individual player ratings on team performance. Specifically, it may be of interest to consider whether the distribution of performances across the 22 players in each team has an effect on team performance.

In team sports, the analysis of objective performance data relating to discrete player actions (i.e., kicks/handballs, whilst factoring in contextual information such as pressure from opponents, field position, time of the match, etc) can be a viable strategic resource. Specifically within AFL teams, objective ratings systems can be used for various aspects of organisational decision support. For example, each AFL club has approximately 45 players on their roster (maximum 47) and is constrained in their ability to recruit players by a salary cap. Furthermore, only 22 of these players are selected to play each round. This in turn puts a greater emphasis on decisions made with respect to player contracting and the development of players within their roster, as well as weekly player selection, respectively. Applications of the AFL Player Ratings could be made in order to gain a greater understanding of what makes an individual player unique, what areas they lack in and also to forecast the level of performance expected from players in the future.

Despite the strength of the PART model produced in phase two of this study, its generalisability is unknown, as it was limited to the 2016 AFL season due to the data availability. In order to test the generalisability of this model, an external validation should be undertaken when data becomes available for subsequent AFL seasons, to assess whether longitudinal variations exist.

Conclusion

The results from this study support the validity of the AFL Player Ratings system and its ability to objectively assess combined player performance in AF. By utilising objective outcomes as dependent variables, a more thorough understanding of how equity is used as a quantifiable measure to relate to successful performance can be achieved. To further refine the generalisability of the model produced in phase two, subsequent seasons of data could be added once they become available. Future work should focus on the continual development of improving the ratings system as new technologies become available, as well as the interpretation and application of the AFL Player Ratings system for objective performance analysis and operational decision-making.

Acknowledgements

The authors would like to acknowledge Champion Data for providing the data used in undertaking this study.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

1. Hughes MD and Bartlett RM. The use of performance indicators in performance analysis. *J Sports Sci.* 2002; 20: 739-54.
2. Travassos B, Davids K, Araújo D and Esteves PT. Performance analysis in team sports: Advances from an Ecological Dynamics approach. *International Journal of Performance Analysis in Sport.* 2013; 13: 83-95.
3. Stefani R and Pollard R. Football rating systems for top-level competition: a critical survey. *Journal of Quantitative Analysis in Sports.* 2007; 3: 1-20.
4. Szczepański Ł. Measuring the effectiveness of strategies and quantifying players' performance in football. *International Journal of Performance Analysis in Sport.* 2008; 8: 55-66.
5. McHale IG, Scarf PA and Folker DE. On the development of a soccer player performance rating system for the English Premier League. *Interfaces.* 2012; 42: 339-51.
6. Radovanović S, Radojičić M, Jeremić V and Savić G. A novel approach in evaluating efficiency of basketball players. *Management.* 2013; 67: 37-45.
7. Streib N, Young SJ and Sokol J. A major league baseball team uses operations research to improve draft preparation. *Interfaces.* 2012; 42: 119-30.
8. Chao-Chien C. Evaluation Model for Baseball Players of National Team-Application of Fuzzy Multiple Criteria Decision Making. *International Research in Education.* 2014; 2: 159-71.

9. Gerrard B. Is the Moneyball approach transferable to complex invasion team sports? *International Journal of Sport Finance*. 2007; 2: 214.
10. Robertson S, Back N and Bartlett JD. Explaining match outcome in elite Australian Rules football using team performance indicators. *J Sports Sci*. 2015: 1-8.
11. Hutchins B. Tales of the digital sublime: Tracing the relationship between big data and professional sport. *International Journal of Research into New Media Technologies*. 2016; 22: 494-509.
12. Ayres I. *Super Crunchers: how anything can be predicted*. Hachette UK, 2008.
13. Norman DA. *Things that make us smart: Defending human attributes in the age of the machine*. Basic Books, 1993.
14. Heasman J, Dawson B, Berry J and Stewart G. Development and validation of a player impact ranking system in Australian football. *International Journal of Performance Analysis in Sport*. 2008; 8: 156-71.
15. Lewis M. *Moneyball: The art of winning an unfair game*. WW Norton & Company, 2004.
16. Stewart M, Mitchell H and Stavros C. Moneyball applied: Econometrics and the identification and recruitment of elite Australian footballers. *International Journal of Sport Finance*. 2007; 2: 231-48.
17. HeraldSun. SuperCoach 2016: Champion Data explains key stats in scoring formula. *Herald Sun*. 2016.
18. Jackson K. Football numbers man brings players to account. In: Marino M, (ed.). Swinburne Venture Magazine: Swinburne University of Technology, 2009.
19. Harrell Jr FE. Package 'Hmisc'. *R Foundation for Statistical Computing*. 2017.
20. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016.
21. Ofoghi B, Zeleznikow J, MacMahon C and Raab M. Data mining in elite sports: A review and a framework. *Measurement in Physical Education and Exercise Science*. 2013; 17: 171-86.
22. Hornik K, Buchta C and Zeileis A. Open-source machine learning: R meets Weka. *Computational Statistics*. 2009; 24: 225-32.
23. Frank E and Witten IH. Generating accurate rule sets without global optimization. 1998.
24. O'Shaughnessy DM. Possession versus position: strategic evaluation in AFL. *J Sports Sci Med*. 2006; 5: 533.