JSHP
JOURNAL OF SPORT AND HUMAN PERFORMANCE

ORIGINAL RESEARCH OPEN ACCESS

# PLAYER RATINGS AND ONLINE REPUTATION IN SUPER RUGBY

Bracewell PJ[1,2], Hilder TA[1], Birch F[3]

[1]*DOT loves data, Wellington, New Zealand*

[2]*Victoria University of Wellington, Wellington, New Zealand*

[3]*Pro-athlete Online, Wellington, New Zealand*

## ABSTRACT

Sports reporting contributes to the entertainment derived from professional sport. Studies have been undertaken that explore sentiment and match outcomes [1,2], but, little is known about the quantitative relationship between on-field performance and mainstream media perception of athletes. This largely stems from robust methods for evaluating on-field performance in a number of sports.

Using commercial tools for assessing rugby player performance (RPI) and rating sentiment (Ethel), this relationship is examined using the five 2019 New Zealand Super Rugby franchise squads from the first round of the 2018 competition until mid-April of 2019. The combination of ratings and current event data generates a summary data set of 2008 observations for analysis.

Simple linear regression is used to test both the statistical significance and inform the interpretation of these findings.

An athlete's playing reputation is derived from a string of on-field performances. This is essentially an estimation of their ability as described by Bracewell [3]. When matches are previewed, this playing reputation informs the number of articles featuring an athlete and the associated sentiment. That is, players perceived to have better ability are talked about more often and more positively. Performances within a match appear to influence the media post-match review. That is, athletes who performed well in a game are more likely to be mentioned and talked about favourably.

Arguably, this is a trivial result as it is no surprise that the "stars" get more coverage as it is more likely to coincide with public interest. However, the ability to quantify and find statistical evidence of these relationships is non-trivial and has important implications for management of athlete reputation. That is, ability influences previews, performances influence reviews. As this can be benchmarked, other influences can be examined to assess how individuals cope with public scrutiny, or identify individuals who's associated ability to generate positive content is rising or falling. This empowers athletes and their agents to identify suitable opportunities with demonstrable reach.

**Keywords:** Player ratings, Mainstream media, Sentiment analysis, Online reputation, Natural language processing

## BACKGROUND

Sports fans are not short of an opinion about their favourite team and players. Arguably, this can be shaped by myriad influences, including player performance and media descriptions. The media has a substantial role in describing events for the purpose of informing and entertaining. Previous work in team sports, such as rugby [1] and cricket [2] has demonstrated that the sentiment and themes within written commentary is statistically significantly associated with match outcomes and in-game events.

However, there is very little research on the association between mainstream media and individual player performance. Knowledge of this association would be useful for reputation management and player welfare, especially from a mental health perspective.

In the context of online media, primary news sources are still a major contributor to the formation of opinion [4], are considered more trustworthy [5] and have a strong impact on people's Twitter sentiment [6]. This suggests that curated digital mainstream media can be considered an important source of an area's sentiment. If correct, then this could provide a new large source of easily accessible data on sentiment. Sentiment analysis is a family of text mining techniques that assign polarity scores to natural language. Typically, it is treated as a supervised machine learning problem. Example sentences are supplied that have been labelled as "positive" and "negative". Given sufficient training data, learning algorithms are able to distinguish positive from negative language. Positive language use scores above 0.0, and negative language scores below 0.0. Importantly digital delivery of news reporting and sports commentary provides a wealth of accurately time-stamped textual data that can be easily indexed via technological means. DOT loves data, a data science firm based in Wellington, New Zealand, has created a transitional index of approximately 25 million news articles, providing us with the ability to create time series of sentiment by author and by topic for the purpose of summarising current events.

Arguably, rugby union is the national sport of New Zealand and the national representative team, the All Blacks, have dominated the international game. Within New Zealand, rugby is widely reported with both mainstream news media and specialist publications.

Collection of rugby player statistics via notational analysis, assignment of opinion-based ratings and attempts to quantify rugby union player contribution are not new [3,7-11]. However, this performance data is not easily accessible. Historically, this stems from the effort required to undertake notational analysis which requires video footage and incurs costs which are invariably passed onto interested parties. For example OPTA and Sports Analytics are commercial supplier s of rugby union data. However, Rugby Pass (https://index.rugbypass.com/) have recently deployed a novel individual playing system [12] which makes this information freely available on their website.

By combining the sentiment and noise about Super Rugby Players in the days around a match, we are seeking to understand the association between media reports and the actual match outcomes for individuals.

## LITERATURE REVIEW

Media is often described as the fourth estate or fourth power in a country. This label given to "The Media" indicates the power and influence people perceive the media to have. The influence of the media on our lives has been of interest for generations [13, 14]. Understanding how the media impacts our lives is important as it can have significant impact on our self-concept [15,16], businesses [17,18], human development [19] and our behaviour [20]. This study aims to contribute to this valuable area of research, by looking at the sentiment within digital mainstream media articles and the association with individual rugby player performance.

Sentiment analysis quantifies the emotional polarity of natural language. Using this quantified sentiment, the respective opinions of the writer surrounding a given topic can be identified. In sport, it is assumed that writers will not all share the same opinion and in fact, some writers pre-match sentiment will be more correlated with events. For instance, McIvor et al [2] demonstrated the potential to predict in-game events in cricket using natural language processing. Specifically, they found that positive comments regarding batters and negative comments regarding bowlers are correlated with more runs being scored. Consequently, that study validated the hypothesis that commentators are storytellers with latent game understanding. Extending this concept further, commentary is important to sports broadcasting as it aims to provide information about the game, such as describing the actions of the players and providing statistics. Duncan & Hasbrook [21] stated that commentary has the effect of drawing the reader's attention to the part of the picture that merit closer attention, an effect called italicized. The commentator is a storyteller that guides the viewers and readers attention to the most important aspects of the game or a given play. Farrell [22] stated that commentators, whether written or spoken, convey in-game action as it happens with accuracy, clarity and concision. Commentators must fulfil the key purpose of the reporting: describing the game objectively, fluently, largely error-free and in a timely fashion.

The emergence of live text commentary has produced a huge amount of text commentary data, but there are few studies about utilising this rich data source. However, the literature primarily focuses on commentary synthesis (often post-game summaries) and analysis of fan emotions via social media. Sport-fan comments posted on social media such as Twitter have been found to contain valuable information regarding in-game events, anticipation of results and fan-base emotion towards results [23,24]. Other studies that utilise text commentary include: [25-29].

Shiladitya et. al. [30] used post-match sentiment analysis gathered from Twitter to predict NFL match results in future games. The sentiment analysis was used within a logistic regression to predict match outcomes and sports betting outcomes. That study showed simple features derived from many tweets (42 million tweets gathered during the 2010-2012 season) can outperform or at least match that of traditional features which use game statistics.

Sport has the benefit of measuring outcomes. Other industries and use-cases may not have objective data for comparison. Turney [31] carried out sentiment analysis on reviews across the film industry, vacation destinations and other areas to create a "thumbs up" or "thumbs down" system via a simple unsupervised classification algorithm. After classifying the reviews, the author found that movie reviews were very hard to classify.

Lucas et al. [32] explored the emotions of the public using sentiment analysis of Twitter posts during the 2014 FIFA World Cup. The purpose of that research was to identify exactly what makes a football match exciting. Their research found that the rate of tweets per minute increased with the difference in scores (one team wins by a significant margin) with 99% confidence. A sentiment analysis was also carried out in which the authors categorised each tweet as: positive, neutral or negative. Research found that, as the rate of tweets increase, the proportion of tweets which are negative increases with a correlation of 0.26. The percentage of neutral and positive tweets showed no relationship with the volume of tweets. This indicates that people tweet more negatively and more often when a team is beaten by a significant margin.

Opinion has been widely analysis outside of sport. Li and Wu [33] used text mining and sentiment analysis for online forums hotspots detection and forecasting. K-means clustering and SVM machine learning techniques were used to identify forum hotspots. Both methods proved to be highly predictive with both methods concluding with the same top 4 hotspot forums. The purpose of this research was to aid the decision-making process for Internet social network users in detecting hotspots.

More traditional source than social media are also fruitful sources of sentiment data. Zhai et. al. [34] used sentiment analysis on headlines of New York Times articles to predict the daily market trend. The classifier was inaccurate on many occasions and often predicted the inverse to what happened. Godbole et. al. [35] carried out large scale sentiment analysis of news articles and blogs. The authors analysed the sentiment surrounding the following topics: business, crime, health, media, politics and sport. They found that sports people often blog positivity most often. It was also found that American politicians speak positively in blogs, but negatively in news articles. Ljajic et. al. [28] used specialized sports dictionaries with hard-coded weights established beforehand. Sentiment polarity could best be classified using a logarithmic difference of ±2 when comparing counts of positive and negative words, where the difference in term frequency multiplied by inverse document frequency for positive over negative terms was logged (DifLogt). DifLogt > 2 implies a positive term classification.

Here, we will evaluate the relationship between the sentiment of pre-match reports from mainstream media and compare against the actual match outcome for individual players. The intent is to understand if this sentiment is predictive of outcome, and if there is meaningful, collective consistency of sentiment across journalists. Finally, the ability to evaluate the post-match insights supplied by journalists is investigated with respect to recent on-field performance.

However, as the intent of this research is to identify and demonstrate any association between onfield performance and online reactions in mainstream media, a challenge arises with the ability to independently access robust player performance data. We were able to source proprietary data from RugbyPass (https://index.rugbypass.com/) who developed a result-driven player rating system for rugby union [12,36], using methodology outlined by Patel et. al. [37] for deriving cricket player ratings. Random forest regression (combined with expert domain knowledge) is used to identify the most important player attributes for each position on a rugby team. These player attributes are derived from commercially acquired data arising from notational analysis [11]. A dynamic probability of winning score is used as the target for the regression model, as binary match outcome is too coarse to be an effective target. This probability of winning is derived by training an ensemble of logistic models to predict match outcomes from time-dependent match statistics. In addition to its role as a target, the probability of winning model enables each player's accumulated influence on probable match outcomes to be directly quantified.

To develop the rating, player features are aggregated to match level and normalised by the minutes played. These features are then grouped by position. For each position a linear transformation of the aggregated information is learnt which sets the variance of each feature to unity. This enables cross-position comparison, and is similar to the process that Bracewell [3] deployed using factor analysis. This is further extended to form an overall rating as a linear combination of features, where the weights in the linear combination are defined by position and are validated by a consensus of rugby experts. To derive player ratings based on a series of matches, their current rating is calculated as an exponentially weighted mean of past ratings. For each game, the match rating and EMWA smoothed rating are obtained.

## DATA SOURCES

Here, the association between the sentiment regarding elite rugby players in NZ and their on-field performances are investigated. This is achieved by using a proprietary natural language processing tool, Ethel, and individual rugby player ratings produced commercially by RugbyPass. Details of these data sources follow.

*Measuing sentiment*

DOT loves data maintains a transitional archive of news content from publicly available sources that it calls the Pressroom which it uses for summarising trends associated with current events. To date, this archive incorporates approximately 25 million articles spanning from 2005 to 2018. The Pressroom contains an index of content produced on online news publications, including a comprehensive collection of articles published on nzherald.co.nz (NZME) and stuff.co.nz (Fairfax NZ). The Pressroom was previously used by Bracewell et al. [38] in conjunction with the Natural Language Toolkit (NLTK) in Python [39] to highlight the sentiment associated with major events. Each article is given a sentiment score between -1 (negative) and 1 (positive).

Once all the data was retrieved, the sentiment of each article was standardized against a historical benchmark. The benchmark sentiment statistics were generated using all articles from the archive spanning from 2013-2018. First, each news item is given a z-score, relative to the mean and standard deviation of the benchmark. The z-score is then normalized using standard normal distribution to provide a p value between 0 and 1. An article is considered 'positive' if its p value is greater than or equal to 0.7. An article is considered 'negative' if its p value is less than or equal to 0.3. Any other article is considered neutral. A sentiment index is then generated which represents the percentage of positive articles minus the percentage of negative articles for a give time period. To compare to the cadence of the Super Rugby season, sentiment index is also determined weekly, before and after the game. Using the kick-off times and publication datetime, we were able to determine if the article was published pre-game or post-game. Tuesday was selected as the end of the week. As a consequence, for each player, articles where they were mentioned were tagged alongside the round and if it was pre-match or post-match. Sample data is shown in Table 1.

*Measuing performance*

As described previously RugbyPass (https://index.rugbypass.com/) has developed a result-driven player rating system for rugby union. DOT loves data were able to access this proprietary information with the permission of RugbyPass. Importantly, player ratings are freely available on that site. These ratings are updated in the hours that follow a Super Rugby fixture.

*Defining the population*

Super Rugby is a professional men's rugby union competition involving teams from New Zealand, South Africa, Japan, Australia and Argentina which was established following the sport becoming professional in August 1996 (https://super.rugby/superrugby/about-super-rugby/). Within New Zealand, players for the five franchises are recruited from the 26 provincial unions, exclusively drawing on talent from the top-tier teams. The 2019 Super Rugby squads were obtained online from each of the five NZ domiciled franchises: www.hurricanes.co.nz/squad/current-squad/, www.theblues.co.nz/players/, www.crusaders.co.nz/the-team, www.thehighlanders.co.nz/our-team/player-profiles, www.chiefs.co.nz/team/team/player-list. These lists were used as the basis of the search string to find associated content and sentiment with the transitional archive used within Ethel. This resulted in a data set with 2008 observations involving 190 players. These players appear at least once resulting in 44,516 articles referencing individual players. Importantly, multiple people may be mentioned in the same article, as shown in Table 1. It is important to note that only data from this time frame, and the week around a match was explored. That is, All Black and Provincial Rugby were not played during this time frame and therefore not in scope.

**Table 1:** Sample sentiment data.

| Date/Time | Article ID | Search Term | URL | Headline | Sentiment |
|---|---|---|---|---|---|
| 14/09/2018 6:55 | #63:2836765 | Kieran Read | https://www.stuff.co.nz/sport/rugby/all-blacks/107106143/all-blacks-take-time-out-from-training-to-play-with-their-children | All Blacks take time out from training to play with their children | 44.8 |
| 14/09/2018 17:00 | #60:2835372 | Beauden Barrett | https://www.nzherald.co.nz/sport/news/article.cfm?c_id=4&objectid=12125506 | Speed demon: Rieko Ioane lays claim to being fastest All Black in history | 69.9 |
| 14/09/2018 17:00 | #60:2835372 | Rieko Ioane | https://www.nzherald.co.nz/sport/news/article.cfm?c_id=4&objectid=12125506 | Speed demon: Rieko Ioane lays | 69.9 |

| | | | | | |
|---|---|---|---|---|---|
| | | | | claim to being fastest All Black in history | |
| 14/09/2018 17:00 | #60:2835372 | Waisake Naholo | https://www.nzherald.co.nz/sport/news/article.cfm?c_id=4&objectid=12125506 | Speed demon: Rieko Ioane lays claim to being fastest All Black in history | 69.9 |
| 14/09/2018 17:00 | #62:2836647 | Ryan Crotty | https://www.stuff.co.nz/sport/opinion/107084376/josh-dugans-tears-ought-to-remind-us-to-be-careful-how-we-criticise-people | Hamish Bidwell: Josh Dugan's tears ought to remind us to be careful how we criticise | 35.3 |

**Table 2:** Summary of models and results testing relationship between sentiment, noise and player ratings.

| Model | | Parameter | Standard Error | t-statistic | p-value |
|---|---|---|---|---|---|
| 1 | sent_pre = match_rating | 0.0222 | 0.0113 | 1.96 | 0.0497 |
| 2 | sent_pre = overall_rating | 0.0537 | 0.0143 | 3.75 | 0.0002 |
| 3 | sent_pst = match_rating | 0.0343 | 0.0124 | 2.77 | 0.0056 |
| 4 | sent_pst = overall_rating | 0.0435 | 0.0157 | 2.76 | 0.0058 |
| 5 | bldup = match_rating | 0.0052 | 0.0012 | 4.20 | <0.0001 |
| 6 | bldup = overall_rating | 0.0104 | 0.0016 | 6.72 | <0.0001 |
| 7 | wrpup = match_rating | 0.0967 | 0.0093 | 10.43 | <0.0001 |
| 8 | wrpup = overall_rating | 0.2162 | 0.0111 | 19.50 | <0.0001 |

## METHODOLOGY

We tested the association between on-field performance based on both a single performance and smoothed across multiple performances, and online reaction, measured as sentiment and the number of articles with at least one mention for all players listed on a 2019 Super Rugby roster. That is, using this season's super rugby squads, we obtained data about their performances and what was written about them across 2018 and 2019.

Simple linear regression was used to examine the association. To satisfy the assumptions of regression, the data required transformation to approximate normality, as shown in composite Figure 1. Match Ratings were left skewed, so cubic and quartic transformations were applied to the match and overall ratings respectively. Sentiment pre and post match was approximately normal. The noise (number of articles) was right skewed. A cubic root introduced approximate normality for pre-match noise, but not post-match noise.

## RESULTS

The intent of the analysis is to demonstrate any relationship between on-field performance and online reactions. We tested 8 simple models which explored concepts relating to noise. Results obtained from the transformed attributes and original attributes were equivalent in the direction of the coefficients and statistical significance at an $\alpha$ of 0.05. As a consequence, for ease of interpretation, the regression coefficients from the untransformed attributes are supplied in Table 2. Sent_pre is sentiment of articles in the period leading into a game (typically 3 to 4 days). Sent_pst is the sentiment from the days immediately following a game (typically 2 to 3 days). Match_Rating is the rating calculated by RPI for that match. This is essentially performance. Overall_Rating is the player's long term rating. This is essentially an indication of ability [3]. Bldup is the number of articles published leading into a match, and Wrpup is the number of articles following the match.

On average, when a NZ Super Rugby Player adds an additional 10 units to their match performance rating which is between 0 and 100, this equates to 1 additional post match media write up (model 7). Simplistically, if a player increases their own team's chances of winning by 10%, this equates to 1 additional media mention. And sentiment most likely to be positive ($r=0.1$, $p<0.01$). Furthermore, sentiment the week before is positively correlated with match performance the week after ($r = 0.1$, $p<0.01$), as is the amount of noise ($r = 0.2$, $p<0.01$). Whilst this is just an association, a crude interpretation is that the media can either influence, or predict, a player's on-field performance.

Reputation has the greatest impact in the number of mentions prior to a match (2.02x) (models 5 and 6). This also follows into post match reviews, where overall playing reputation carries twice as much weight compared to a single performance for post match media mentions (comparing coefficients from models 7 and 8).

Playing reputation is 2.4 times more important on sentiment compared to single performances in match previews (0.053726/0.022197) (comparing coefficients from models 1 and 2). Furthermore, comparing coefficients from models 3 and 4 indicate a one-off performance can match a playing reputation for level of sentiment in post match reports (no statistically significant difference 90% CI [0.00014, 0.00055]).

This methodology can be used to track a player's on-field and online 'reputation', and therefore used to assist reputation risk management of athletes. Figure 2 illustrates the 'reputation' of Otago Highlanders and All Blacks player, Waisake Naholo during his last 20 Super Rugby Appearances. As shown, both Naholo's on-field and online reputation have decreased during this time ($r=0.53$). A tool based on these metrics could be used by, for example, 1) player agents to assist in future career decision making, 2) coaches to identify players who are more affected by what is said off the field, or 3) digital strategists to develop content to increase positive sentiment and drive reputation beyond the player's performance.

**Figure 1.** Histograms of attributes used to test association between on-field performance and reporting

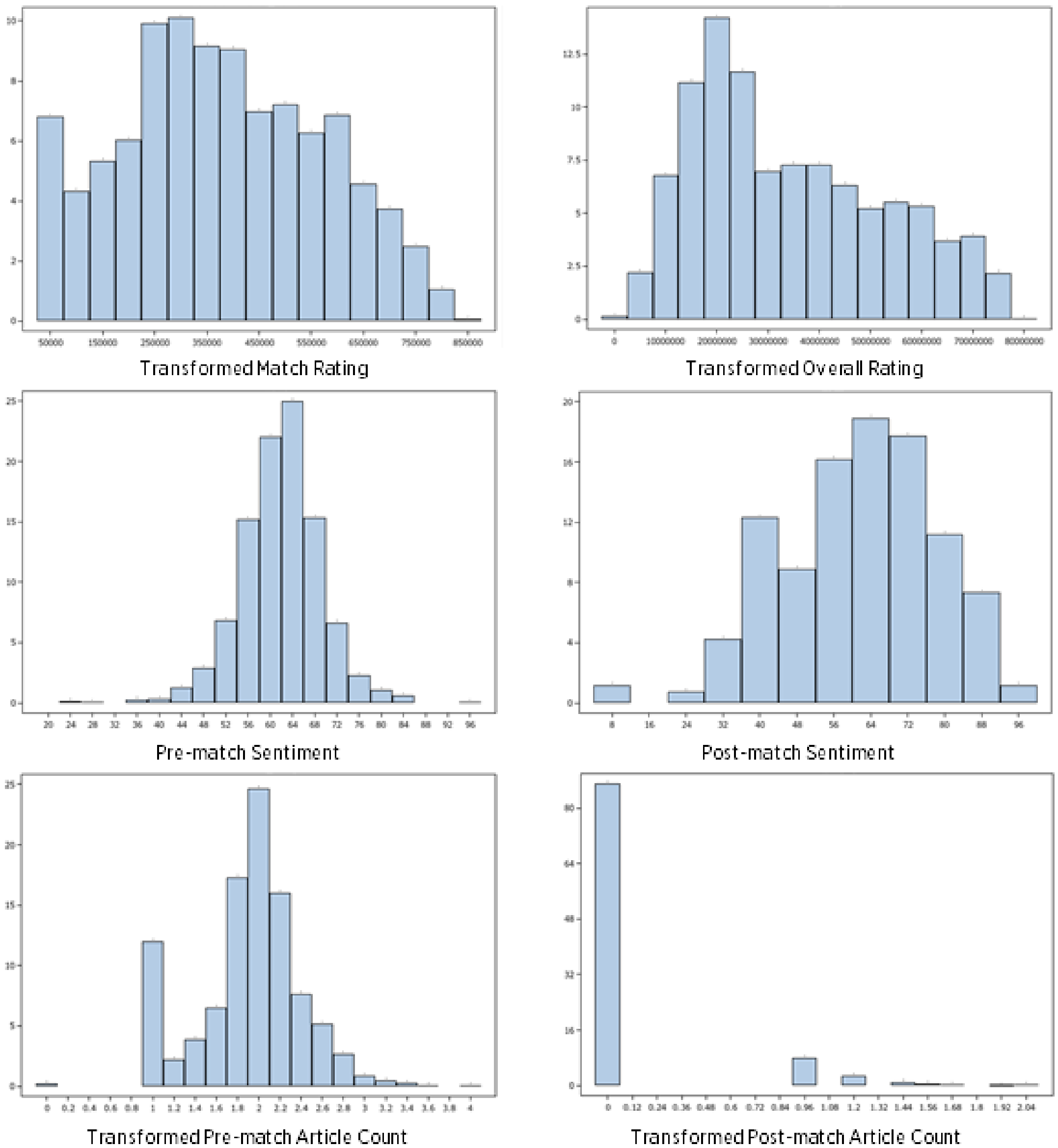
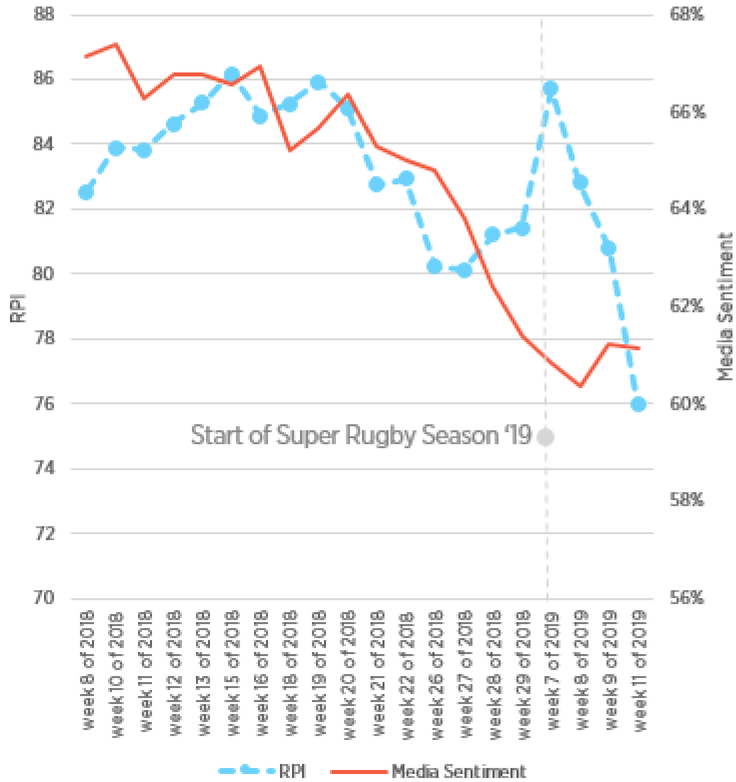



**Figure 2.** Otago Highlander and All Black, Waisake Naholo's 'reputation': on-field and online




## DISCUSSION AND CONCLUSION

We sought to understand the association between media reports and the actual match outcomes for individuals using five 2019 New Zealand Super Rugby franchise squads from the first round of the 2018 competition until mid-April of 2019. We demonstrated that there is a meaningful relationship between on-field performance and online reactions. Importantly, a player's playing reputation has the greatest impact on a player's online profile, in terms of the number of articles and sentiment. Understanding this association means we have a robust, reliable framework to explore further.

Although it is no surprise that "stars" get more coverage, we have demonstrated that on-field performances are associated with metrics extracted from online content. As a consequence, players can shape their reputation by what they do on the field. Similarly, reporters can also influence a player's reputation. Whilst we do not demonstrate cause and effect, the presence of a statistically significant association that aligns with a pragmatic interpretation suggests that there may be wider implications for the interaction between mainstream media and players.

This work illustrates the potential to combine reputation risk management with both sports ratings and natural language processing of mainstream media. Such an approach will enable delivery of a scalable solution for professional athletes and their associates to understand the impact of their on-field and off-field behaviour on their personal brand. This would aid strategic decisions around the type of content to develop, the best timeline to deploy certain content and a measurement tool to assess the impact of that content. Moreover, such a tool could provide the ability to identify mental health risks. For example, the resilience of a player to public scrutiny could help understand which players need more support.

**REFERENCES**


1. Simmonds, PP, McNamara, TS & Bracewell, PJ. Predicting win margins with sentiment analysis in international rugby. *Proceedings of the 14th Australian Conference on Mathematics and Computers in Sports*. Ray Stefani & Anthony Bedford eds. Sunshine Coast, Australia: ANZIAM Mathsport, 2018, pp. 137-142.
2. McIvor, JT, Patel, AK, Hilder, T & Bracewell, PJ. Commentary sentiment as a predictor of in-game events in T20 cricket. *Proceedings of the 14th Australian Conference on Mathematics and Computers in Sports*. Ray Stefani & Anthony Bedford eds. Sunshine Coast, Australia: ANZIAM Mathsport, 2018, pp. 44-49.
3. Bracewell, PJ. Monitoring Meaningful Rugby Ratings. *Journal of Sports Sciences*, 21 (8), 2003, pp. 611-620.
4. Ying, KT, Minhao, Z, Bob, D, Paul, C & Philip, G. Insight from the horsemeat scandal: Exploring the consumers' opinion of tweets toward Tesco. *Industrial Management & Data Systems* 116(6), 2016, pp. 1178-1200.
5. Ceron, A. Internet, news, and political trust: The difference between social media and online media outlets. *Journal of Computer-Mediated Communication* 20(5), 2005, pp. 487-503.
6. Bollen, J, Mao, H & Pepe, A. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *Paper presented at the Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain; 2011.
7. Quarrie, KL & Hopkins, WG. Changes in player characteristics and match activities in Bledisloe Cup rugby union from 1972 to 2004. *J Sports Sci* 25(8), 2007, pp. 895–903.
8. Quarrie, KL, Hopkins, WG, Anthony, MJ & Gill, ND. Positional demands of international rugby union: Evaluation of player actions and movements. *Journal of Sport Science and Medicine in Sport* 16(4), 2013, pp. 353-359.
9. Deutsch, MU, Kearney, GA, & Rehrer, NJ. Time-motion analysis of professional rugby union players during match-play. *J Sports Sci* 25(4), 2007, pp. 461–472.
10. Hughes, M. Notation analysis in football. In Science and Football II (edited by T. Reilly, J. Clarys and A. Stibbe), pp. 151–159; 1993. London: E & FN Spon.
11. Hughes, M. Notational analysis – a mathematical perspective. *International Journal of Performance Analysis in Sport* (4:2), 2004, pp. 97-139.
12. Bracewell, PJ, Moore, WE, McIvor, JT & Stefani,R. Deriving result-driven rugby player performance ratings. *Proceedings of the 14th Australian Conference on Mathematics and Computers in Sports*. Ray Stefani & Anthony Bedford eds. Sunshine Coast, Australia: ANZIAM Mathsport, 2018, pp. 131-136.
13. Rosten, LC. President Roosevelt and the washington correspondents. *The Public Opinion Quarterly* 1(1), 1937, pp. 36-52.
14. Tichenor, PJ, Donohue, GA & Olien, CN. Mass media flow and differential growth in knowledge. *The Public Opinion Quarterly* 34(2), 1970, pp. 159-170.
15. Groesz, LM, Levine, MP & Murnen, SK. The effect of experimental presentation of thin media images on body satisfaction: A meta-analytic review. *International*

*Journal of Eating Disorders* 31(1), 2002, pp. 1-16.
16. Milkie, MA. Social comparisons, reflected appraisals, and mass media: The impact of pervasive beauty images on black and white girls' self-concepts. *Social Psychology Quarterly* 62(2), 1999, pp. 190-210.
17. Fang, L & Peress, J. Media coverage and the cross-section of stock returns. *The Journal of Finance* 64(5), 2009, pp. 2023-2052.
18. Hoffman JM. A framework for understanding the public's perspectives of mining applied to the kentucky coal industry: University of Kentucky; 2003.
19. Villani, S. Impact of media on children and adolescents: A 10-year review of the research. *Journal of the American Academy of Child & Adolescent Psychiatry* 40(4), 2001, pp. 392-401.
20. Anderson, P, de Bruijn, A, Angus, K, Gordon, R & Hastings, G. Impact of alcohol advertising and media exposure on adolescent alcohol use: A systematic review of longitudinal studies. *Alcohol and Alcoholism* 44(3), 2009, pp. 229-243.
21. Duncan, MC & Hasbrook, CA. Denial of power in televised women's sports. *Sociology of sport journal* 5(1), 1988, pp. 1-21.
22. Farrell. Pitch perfect: the fine art of live sports commentary. [2018, May 15]. Retrieved from https://www.rte.ie/eile/brainstorm/2018/0515/963663-pitch-perfect-the-fine-art-of-live-sports-commentary/.
23. Sinha, S, Dyer, C, Gimpel, K & Smith, NA. Predicting the NFL using Twitter. arXiv preprint arXiv:1310.6998; 2013.
24. Schumaker, RP, Labedz Jr, CS, Jarmoszko, AT & Brown, LL. Prediction from regional angst–A study of NFL sentiment in Twitter using technical stock market charting. *Decision Support Systems* 98, 2017, pp. 80-88.
25. Lareau, F, Dras, M & Dale, R. Detecting interesting event sequences for sports reporting. In *Proceedings of the 13th European Workshop on Natural Language Generation* (pp. 200-205). Association for Computational Linguistics; September 2011.
26. Zhang, J, Yao, JG & Wan, X. Towards constructing sports news from live text commentary. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* 1, 2016, pp. 1361-1371.
27. Lee, G, Bulitko, V & Ludvig, EA. Sports Commentary Recommendation System (SCoReS): Machine Learning for Automated Narrative. In AIIDE; October 2012.
28. Ljajić, A, Ljajić, E, Spalević, P, Arsić, B & Vučković, D. Sentiment analysis of textual comments in field of sport. In *24nd International Electrotechnical and Computer Science Conference* (ERK 2015), IEEE, Slovenia; September 2015.
29. Minard, AL, Speranza, M, Magnini, B, Qwaider, MR & Kessler, FB. Semantic interpretation of events in live soccer commentaries. *CLiC it*, 2016, pp. 205.
30. Shiladitya, S, Dyer, C, Gimpel, K & Smith, N. Predicting the NFL Using Twitter. Presented at *ECML/PKDD 2013 Workshop on Machine Learning and Data Mining for Sports Analytics*; 2013.
31. Turney, P. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, July 2002, pp. 417-424.
32. Lucas, G, Gratch, J, Malandrakis, N, Szablowski, E, Fessler, E & Nichols, J. GOAL! Using Sentiment in the World Cup to Explore Theories of Emotion. *Image and Vision Computing* 65, 2014, pp. 58-65.

33. Li, N & Wu, D. Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems* 48(2), 2010, pp. 354-368.
34. Zhai, J, Cohen, N & Atreya, A. Sentiment analysis for news articles for financial signal preparation. S224N/Ling284 Final Projects 2010, 2011, pp. 11.
35. Godbole, N, Srinivasaiah, M & Skiena, S. Large-Scale Sentiment Analysis for News and Blogs. Presented at *International Conference on Weblogs and Social Media*; 2007.
36. Cully, P. The Kiwi data 'genius' who made a model to rate every rugby player in the world. *Stuff*, [2019, April 21] Retrieved from https://www.stuff.co.nz/sport/rugby/all-blacks/112155698/the-kiwi-data-genius-who-made-a-model-to-rate-every-rugby-player-in-the-world.
37. Patel, AK, Bracewell, PJ & Rooney, SJ. An Individual-based Team Rating Method for T20 Cricket. *Journal of Sport and Human Performance* 5(1), 2017, pp. 1-17.
38. Bracewell, PJ, McNamara, TS & Moore, WE. How Rugby Moved the Mood of New Zealand. *Journal of Sport and Human Performance* 4(4), 2016, pp. 1-9.
39. Bird, S, Klein, E & Loper, E. Natural Language Processing with Python. O'Reilly Media; 2009.