

Title: Identification & interpretation of playing styles in the National Football League (NFL).

Ben Carter¹

ORCID: 0009-0009-9706-4505

LinkedIn: www.linkedin.com/in/ben-carter-4a7417114

Twitter: https://x.com/Ben_a_Carter

Department of Health, Exercise Science, and Recreation Management, School of Applied Sciences,
University of Mississippi

Nic James²

ORCID: 0000-0002-4181-9501

LinkedIn: <https://www.linkedin.com/in/nic-james-92b03b29/>

Twitter: <https://x.com/ProfNicJames>

London Sport Institute, Faculty of Science and Technology, Middlesex University

Nimai Parmar³

ORCID: 0000-0001-5540-123X

LinkedIn: <https://www.linkedin.com/in/nimai-parmar-6bb73448/>

Twitter: <https://x.com/NimaiParmar>

London Sport Institute, Faculty of Science and Technology, Middlesex University

Correspondence:

Dr. Ben Carter

Health, Exercise Science, and Recreation Management, School of Applied Sciences, University of
Mississippi

310 Garland Hall, University, Oxford, 38655, Mississippi, USA

+01 (662) 380-9770 or +44 (0)7544 923 485

CarterB@olemiss.edu; B.Carter@mdx.ac.uk; Ben.A.Carter@icloud.com

Funding, disclosures and data:

This research was conducted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) at Middlesex University. No Funding was awarded for this study to be conducted, and the authors report there are no competing interests to declare. The data that support the findings of this study are openly available in nflfastR at <https://www.nflfastR.com> (Sebastian and Baldwin, 2024) and pro-football-reference (2023) at https://www.pro-football-reference.com/about/advanced_stats.htm.

Identification & interpretation of playing styles in the National Football League (NFL).

Abstract

The ability to predict opposition behaviours informs critical decision-making. Moreover, understanding differences in offensive and defensive patterns of play has emerged as a limitation in American Football. However, an absence of data-driven foundations from which to understand areas of play confound current perceptions of team interactions. Therefore, the aim of this study was to derive and label playing styles in the NFL. Principal component analysis was used to explore underlying structures responsible for correlations among variables over a period of five seasons (2018-2022, 2622 team observations). Optimal coordinates extraction derived 8 distinct features accounting for 57% of the total variance within the dataset. Direct oblimin (oblique) rotation identified a low, yet meaningful positive correlation between the short and long passing latent variables (.35). Offensively, yards before and after the catch or contact alongside third down success were associated with passing, rushing, and scrambling-related playing styles. Including exclusively defensive performance variables led to 3 factors interpreted as **coverage, disruptions and tackling. These methods provide a means by which team performance may be further analysed, categorized, and potentially aid in predicting and rationalising play design.**

1 Introduction

1.1 Rationale

Advancements in granularity and availability of data collected in sport continues to enhance investigations - increasing the scope of potential research questions (Han, Pei and Kamber, 2011). However, unlike within analogous sports, playbook-centred playing styles (e.g., west-coast offense), rather than data-driven styles of play (Lago-Peñas et al., 2017) persist as means of interpreting teams' behavioural patterns (Teramoto et al., 2015). Notwithstanding, a focus of contemporary analyses of American football have been aimed at understanding the outcomes of decisions in relation to match outcome and formulating novel variables to assess likelihood of drive and performance success (Biro and Walker, 2022). Dissimilar to alternative football codes such as association football, rugby union or league, differences in game mechanics exist wherein possessions are comprised of individual plays interrupted for a period of 40 seconds. Therefore, researchers have sought to increase prediction accuracy by quantifying scoring or win potential using metrics including Expected Points (Carter and Machol, 1971), Win Probability (Lock and Nettleton, 2014), Wins Above Replacement (Yurko, Ventura and Horowitz, 2019). Albeit, in practice, these methods serve predominantly as a retrospective means of contextualising the influence of on-field actions, developing a statistical approach to the prediction of behaviours in American football may enhance the rational underpinning play-calling in the NFL (Emara et al., 2017; Joash Fernandes et al., 2020; Lee, Chen and Lakshman, 2016; Teich, Lutz and Kassarnig, 2016).

Throughout the National Football League, understanding the factors contributing to the predictability of a team's actions provides coaches with actionable insights about the likelihood of collective opposition behaviours (Joash Fernandes et al., 2020). Whereas previous attempts

to evaluate optimal tactical approaches to play-calling have compared teams' proficiencies in outperforming their opponents in either passing or running efficiency (May et al., 2010), such research designs, using outcome data (e.g., total passing/rushing yards, yards per pass/rush), may be limited in their predictive capacity, and as indicators of team tendencies. However, evaluating propensities in play type (i.e., passing or running game) has been the subject of several predictive models (Lee et al., 2016; Emara et al., 2017; Joash Fernandes et al., 2020; Goyal, 2020), predominantly attempting to incorporate situational variables (e.g., down, yards to go, match status, points differential, location and game clock), in addition to **success-related team profile** characteristics over the duration of a season (e.g., pass/run attempts %, pass completion %). Notwithstanding, these algorithms have demonstrated substantial inter-performance and inter-season variability (Lee et al., 2016).

Unprecedented and novel observations of novel microscopic behaviours have been enabled by increasingly comprehensive and abundant data (Jeon and Park, 2021). However, sports such as American football have limited research in performance analysis (Cohea and Payton, 2011) with a consequent paucity of substantiated theory regarding relationships between performance variables (Zuccolotto, Manisera and Sandri, 2018). Moreover, the exploration of complex behavioural data could be used to determine styles of play - hypothesised to account for the variability in play type prediction accuracy (Joash Fernandes et al., 2020). **Indeed, how teams play is influenced by the interrelationships between performance variables (Greenham, Hewitt and Norton, 2017). More specifically,** playing styles are a manifestation of patterns in performance, systematically or tactically implemented by teams over the duration of matchups, competitions and seasons (Fernandez-Navarro et al., 2018; Gollan, Ferrar and Norton, 2018). In equivalent football codes, styles of play have been established through the empirical reduction of indicators into latent variables (Decroos and Davis, 2020; Fernandez-Navarro et al., 2016; Gómez et al., 2018; Kong et al., 2022; Lago-

Peñas et al., 2017; Ruan et al., 2021; Sheenan et al., 2022; Wedding et al., 2021), an objective technique to derive underlying structures in data or correlated areas of performance. Teams' relative dependence or proclivities towards these areas of play can be sustained as a profile from which teams may aim to deviate from or adapt to in accordance with contextual demands of a match and their opponent. Whereas examining differences in these features of offensive and defensive performance between teams has been highlighted as a limitation in American Football (Fokoue and Foehrenbach, 2013; Joash Fernandes et al., 2020), it has been a recent objective in association football - projected to assist in developing a comprehensive overview of play style dynamics (Lago-Penas et al., 2017; Gomez et al., 2018; Gollan et al., 2020). Nonetheless, personnel, formation and positional permutations have previously been used to categorise NFL teams according to playing styles and quantify relative injury risks (Teramoto, Petron, Cross and Willick, 2015). However, these styles of play were ascribed to teams according to perceived similarity with somewhat antiquated systems of play such as the West Coast or Air Coryell offences, as well as the 3-4 or 4-3 defences formulated by coaches in the 1960s and 1970s (Carroll, 1999). Though Teramoto et al. (2015) were able to contextualise their findings through a 'philosophical' framework, reflective of researcher interpretation of play designs (e.g., short or long pass attacks, aggressive and preventative defences), they were unable to empirically account for the game styles present in NFL performance.

Further challenges have been elucidated in the analysis of American football performance, namely in study design wherein defensive play is afforded proportionally scant consideration among action variables, overlooking apparent interactions between teams and potentially omitting half of the events during a match (Fokoué and Foehrenbach, 2013). The effect has been associated with a tendency to regard the dynamics of American football as though the defence is considered constrained to reactionary behaviours and the offence dictates play through possession of the ball (Joash Fernandes et al., 2020). Nonetheless, while it may

be inferred that defensive and offensive game styles are subject to team interactions, offensive and defensive play-by-play metrics collected in the NFL are limited to observations of direct player involvements (Dutta, Yurko and Ventura, 2020). Cohea and Payton (2011) assessed 67 variables, reduced into 14 features of performance, with the aim of determining their association with success. However, this research design demonstrated limited practical applicability as the sampled variables were quantified in terms of outcomes without consideration of the actions occurring during the plays. Additional constraints of granularity for within play measurements has been demonstrated in the capacity of models to correctly predict scrambling attempts. For example, Lee, Chen and Lakshman (2016) identified that a substantial number of mistakes developed in their algorithm concerned teams with ‘mobile’ quarterbacks. These errors, amounting to 77% of occasions, originated on plays that were designed for passing attempts and ended with the quarterback running the ball.

Several approaches to the depiction of playing styles in team invasion sports stem from variation among input variables and their respective data structures (Plakias et al., 2023). This has resulted in attempts to derive distinct phases of play over the duration of possessions using spatiotemporal information to identify combinations of events such as ball movements, positional progressions and player involvement in sequences (Amatria et al., 2021; Cintia and Pappalardo, 2021; Clijmans, Van Roy and Davis, 2022). Findings from such approaches in association football have established patterns of play and elucidated strategies (e.g., elaborate and direct buildup). However, the classification of behaviours using spatiotemporal data in the NFL can be measured over substantially shorter intervals, research has described elements of offensive and defensive play including route running trajectories, blocking vectors and receiver separation (Deshpande and Evans, 2020; Dutta, Yurko and Ventura., 2020; Haar et al., 2019; Ibrahim, 2021; Yurko et al., 2020; Nguyen, Yurko and Matthews, 2023). Albeit, while in analogous sports, individuals are able to implement or re-initiate a coach’s game style

throughout sequences, comparable plays in American football are coordinated - allowing for a systematic formulation of a response to contextual factors (Joash Fernandes et al., 2020). However, once plays have been implemented, post-snap (ball in play) actions quickly become unique with individuals reacting dynamically according to their respective or irrespective of attacking and defensive responsibilities (Dutta et al., 2020). Moreover, these methods often rely on the accessibility of GPS (Global Positioning System) or LPS (Local Positioning System) measurement of movements, presenting an impracticality in data collection due to limitations in team willingness to share resources (Lord, Pyne, Welvaert and Mara, 2020). Furthermore, according to Lord et al. (2020), the complexity of spatiotemporal metrics renders the applicability and ease of interpretation challenging for coaching purposes.

Contrarily, the simplicity and universal familiarity with technical variables allows for strategy and tactics to be ratified according to context (Lord et al. 2020). Notwithstanding, exclusive use of unidimensional performance variables has been characterised as inadequate regarding its sensitivity for identifying playing styles in team sports (Wedding et al., 2021). To accommodate for this effect, research on alternative football codes, namely association football, have engaged numerous methods to discern styles play (Plakias et al., 2023). The most notable and prevalent among these techniques is Principal Components Analysis (Gløersen et al., 2018; Parmar et al., 2018; Fernandez-Navarro et al., 2016; Lago-Penas et al., 2018; Winter & Pfeiffer, 2016; Gollan et al., 2018), an unsupervised approach to dimensionality reduction found to produce effective results in prediction accuracy, efficiency in computational time and validity in reduction of features (Zebari et al., 2020). Therefore, to understand styles of play and their respective compositions, technical data, used in conjunction with dimensionality reduction techniques, may produce an objective and easily interpretable exploration of playing styles in professional American football.

1.2 Objectives

The aim of this study was to address the hypothesised presence of playing styles in the NFL and American football, whilst accounting for a greater degree of granularity than previous investigations. Moreover, this research design sampled dynamic offensive and defensive technical performance variables - incorporating several exclusively defensive variables to avoid offensive bias in game style interpretation. Furthermore, this study sought to determine the strength of relationships between resultant styles of play.

2. Method

2.1 Sample

Game data collated from a play-by-play dataset comprising of 1311 regular season performances from the 2018 to 2022 seasons of the National Football League resulted in 2622 team observations. The data was obtained from third-party public information sources. These resources included nflfastR (2023) and pro-football-reference (2023) - repositories of scraped first-party official game summaries published by the NFL (2023). The original text-based data was comprised of raw play-by-play descriptions of on-field events and actions. Data scraping techniques were exercised on NFL game summaries to generate performance and situational game variables (Appendix A). Whereas certain variables published by nflfastR (Sebastian and Baldwin, 2024) were made available through the nflreadR (Ho and Carl, 2023) package for R, the remaining proportion of the data was ethically web-scraped from pro-football-reference (2023) on the public domain.

2.2 Procedure

Potential performance and situational indicators of playing style ($n = 46$) were collated from the nflfastR (Sebastian and Baldwin, 2024) and pro-football-reference datasets, aggregated, then statistically analysed using the RStudio Integrated Development Environment (IDE) for R (Posit team, 2023), powered by the statistical computing language R (version 4.0.5, R Core Team, 2023). Operational definitions were elucidated through the appraisal of definitions supplied by third parties in accordance the NFL official glossary of terms (Appendix A). Ethical approval for this study was granted by the London Sport Institute Research Ethics Committee (REC) on 24/11/2021 (Appendix B).

2.3 Statistical analysis

Based on a correlation matrix, principal component analysis (PCA) was used to reduce the dimensionality and effectively summarise the correlations between the original variables, whilst minimising any potential loss of variation in the data (Jolliffe, 2002; Tabachnick and Fidell, 2012). The suitability of the data for dimensionality reduction techniques was assessed by the Kaiser-Meyer-Olkin (KMO) test of sampling adequacy (Kaiser, 1974), in addition to Bartlett's (1954) test of sphericity. Two non-graphical criteria, parallel analysis (Horn, 1965), simulating sampling error and optimal coordinates (Raïche et al., 2006), providing an objective interpretation of Cattell's (1966) scree plot were used to gauge the appropriate number of components to retain for analysis (Figure 1). Furthermore, sampling adequacy measured using the Kaiser-Meyer-Olkin (KMO) test resulted in an overall value of 0.74 (>0.7) - classified as middling (Kaiser, 1974). Bartlett's (1954) test of sphericity demonstrated that the data was suitable ($\chi^2 = 1248.3$, $p < 0.001$, $df = 46$) for exploratory factor analysis (EFA). Rotation methods were employed to maximise factor loadings within individual components, whilst minimising the corresponding variable in others to optimise the interpretability of principal components (Plonsky, 2015). However, as oblique rotation solutions should be explored in

cases where rationale exists to hypothesise a relationship between components (Field, 2007) such as teams' proclivity for playing styles relative to others, therefore, the components were rotated using Direct Oblimin. Comrey and Lee's (1992) classifications of contribution quality were used to evaluate variables' association to the identified PCs - excluding variables with loadings lower than $|0.32|$ - accounting for approximately 10% of the variance within a component.

3 Results

[Figure 1 approximately here]

Parallel analysis in the scree test for non-graphical solutions identified 10 distinct areas of play or playing styles (Figure 1), whereas, optimal coordinates, determined by a series of linear equations quantifying the difference between observed and expected eigenvalues demonstrated a clearer delineation ($n = 8$) between meaningful and unambiguous components (Ruscio and Roche, 2012). These factors accounted for 57% of the variance present within the match data (Table 1), while the 10 features determined by the parallel analysis explained a combined 2% additional variance.

[Table 1 approximately here]

The first principal component (TC2) accounting for the greatest proportion of variance present among the match variables (12%) was comprised of predominantly passing efficacy related performance variables. This extensive passing feature included 5 variables with 'excellent' loadings or values $>|0.72|$ (completed passes, passing first downs, passing yards after catch, short passes and passing yards) and 2 variables providing good contributions $>|0.55|$ (pass attempts and quarterback drops). Lesser contributory performance variables achieved fair

loadings $>|0.45|$ (successful passing third downs and receiving tackles broken), in addition to a negative, yet poor association ($>|0.32|$) with combined defensive tackles (Table 2). A secondary passing factor (TC2) accounted for 9% of the variance in the PCA. However, contrary to TC2 with excellent contributions from short passes, 3 of 5 variables (deep passes, intended passing air yards, and passing air yards completed) constituting TC6 demonstrated particularly strong associations within the feature $>|0.9|$. Furthermore, these variables and excellent loadings were rotated with inaccurate and incomplete passes displaying good and fair values respectively.

[Table 2 approximately here]

The third principal component explaining 8% of the variance extant within the data incorporated 5 scrambling variables. Scramble yards, attempts and first downs produced the greatest contributions ($>|0.9|$) towards the factor (TC3). Additional variables concerning scrambling on third down attempts showed reduced importance to the component, albeit with disparate factor scores. The rotation resulted in contribution values classifying successful scrambling third downs as excellent, whereas unsuccessful scrambling third downs was represented as poor. Seemingly unrelated to the passing game, TC4 - accounting for 8% of proportional variance was comprised of rushing variables. Rushing yards, first downs, attempts, yards before represented the running game feature with excellent factor loadings. While they did not meet Comrey and Lee's (1992) highest criterion ($>|0.72|$), rushing yards after contact, successful and unsuccessful third downs provided a very good contribution. Further aspects constituting the rushing game style were tackles broken on running plays (0.38), while combined defensive tackles were negatively associated with the rushing component to a similar degree as TC2.

[Figure 2 approximately here]

TC1 exhibited associations between passing and receiving related indicators - forming 6% of the variance within the unreduced data. Unsuccessful passing third downs (0.81) manifested as its greatest constituent; however, incomplete passes and quarterback hits were associated with TC1 with good (0.58) and fair (0.49) associations respectively. This feature was further loaded with meaningful, albeit poor ($>|0.32|$) contributions from inaccurate passes (0.42), unsuccessful passing fourth downs (0.39), short passes (0.39), pass attempts (0.38), quarterback drops (0.38), receiving dropped passes (0.35). Though only explaining 3% of the total variance, broken tackles during rushing attempts (0.6) and defensive missed tackles (0.58) were established as the most meaningful variable in TC8. However, defensive hurries and receiving tackles broken produced lesser (or poor) factor scores.

The primary defensive component aggregated defensive completed air yards and yards after the catch conceded - enhancing the feature with very good loading values. Moreover, defensive combined tackles were associated poorly with the principal component (TC5). Notwithstanding, ancillary inputs demonstrating intent to reaching the quarterback including defensive quarterback knockdowns (0.49) and defensive blitzes (0.39) were represented within this latent variable. The second underlying defensive structure (TC7) retained variables associated with **disruption**. Most notable among these defensive behaviours - sacks (0.61), tackles for loss (0.51), and interceptions (0.47). Furthermore, TC7 these defensive **disruptions** were substantiated by the positive association of defensive quarterback knockdowns, blitzes, and hurries as well as the negative loading of unsuccessful passing fourth downs exceeding the $|0.32|$ factor score criterion.

Oblique rotation (direct oblimin) conducted on the extracted features revealed the presence of one meaningful positive correlation (.35) between two passing features (Table 5). This correlation demonstrated a low effect between efficacy on short passing patterns in play (TC2) and the deep pass feature (TC6). Though the remaining factors were considered uncorrelated, negative relationships were observed between the running style of play (TC4) and passing related factors (TC2, TC6, TC1).

[Table 3 approximately here]

4 Discussion

Through the reduction of performance variables, this study confirmed the presence of offensive ($n = 5$) and defensive ($n = 3$) playing styles hypothesised to exist in professional American football (Joash Fernandes et al., 2020) between the 2018 and 2022 National Football League seasons. These principal components were derived from 46 performance variables and explained a majority (57%) of the variance presence within the 2622 match observations - allowing for the identification and description of patterns in play reflecting discrete playing styles (Rein & Memmert, 2016). Regardless of feature rotation, certain performance indicators, outlined as marker variables, represent purer and exclusive measurements of factors relative to others (Tabachnick and Fidell, 2012). Furthermore, marker variables can be used to clearly define the nature of a factor and decrease ambiguity in their interpretation. The findings of the principal components analysis revealed three components related to the passing performance, two factors representing running performance and three defensive features.

Two passing components (TC2 and TC6) accounted for the greatest proportion of variance within the extracted features and data yet demonstrated clear differences in patterns of play. TC2 was characterised by marker variables including completed passes, passing first

downs, passings yards after the catch, passing yards and successful passing first downs - performance indicators associated with passing efficacy. Whereas TC2 included playing style characteristics including short passes (<15 yards), quarterback drops and receiving tackles broken, TC6 presented with pure measures of deep **passing** (>15 yards), intended air yards and passing air yards completed, in addition to inaccurate and incomplete passes, albeit representing non-marker variables. Despite recording metrics including calculated measures and their composite variables (e.g., offensive passing completions, offensive passing attempts and offensive passing completion %) with inherent correlations, Cohea and Payton (2011) identified a factor labelled as 'offensive passing and total offence' including passing yards, first downs and completions which exhibited an association with passing efficacy. However, the present analysis enables a greater illustration of the patterns in play influencing passing performance - distinguishing two novel components within passing offence and delineating variables associated with short passing and efficacy (TC2), and long passing and accuracy (TC6). Moreover, these two passing features accounting for 38% of the proportional variance were associated with one another, resulting in a low positive correlation (.35). Meanwhile, an uncorrelated third passing related factor marked by unsuccessful passing third and fourth downs, quarterback hits, and receiving dropped passes was established (TC1). With additional contributions to the feature including unsuccessful passing fourth downs as well as several performance indicators from the other passing styles of play. TC1 exemplifies a relationship between passing in pressing scenarios and the inability to complete passes. Indeed, as the likelihood of a pass play occurring supersedes alternatives on third (79.3%) and fourth (67%) down (Joash Fernandes et al., 2020), the predictability of play type selection appears to reduce the capacity for the passer and the receiver to create separation from defensive personnel and make a catch (Hear, 2019) - resulting in a feature representing **receiving and passing efficiency**.

Quarterback scrambles typically occur on failed passing plays resulting in an unplanned and unforeseeable event (Reyers and Swartz, 2023). Yet, despite findings suggesting that quarterback rushing and scrambling ability is a key determinant of NFL success for college players, Craig and Winchester (2021) discovered that scrambling ability is systematically undervalued and overlooked throughout the scouting process. However, this study, incorporating scrambling related indicators in exploratory factor analysis, has revealed a feature exhibiting 5 marker variables denoting total scrambling yards, attempts, first downs achieved and success on third down as pure measures of scrambling or quarterback rushing ability (TC3). Similarly, the rushing factor (TC4) produced 7 marker variables. Moreover, yards achieved before and after contact exhibited a greater association with rushing performance than success on third down attempts - further corroborated by a negative correlation with opponent defensive combined tackles and a positive correlation with rushing tackles broken, albeit contributing 'poor' factor loadings. Though findings by Walsh (2021) accounted for avoided tackles and yards after contact among other action variables through an overlap of extraction techniques designed to describe running back positional profiles, rushing yards before contact provides a novel and empirically established characteristic of rushing performance in the NFL.

Many quantifications of professional American football performance omit defensive analysis, fewer still observe uniquely defensive measures (Dutta et al., 2020; Walsh, 2021). In cases where defensive actions are represented within a sample, they are often simply parallels of offensive metrics (Cohea and Payton, 2011). While there may be a disassociation between offensive and defensive intra-team players, defensive interactions have been anecdotally construed as reactionary - inhibiting teams' capacity to impose a defensive style (Joash Fernandes et al., 2020). However, this study elucidates the presence of 3 defensive styles of play accounting for 23% of variance explained by the retained principal components. TC5 was

associated with **two coverage related** marker variables comprising of completed air yards and yards after the catch conceded alongside lesser contributions from defensive quarterback knockdowns, combined tackles and blitzes which may be interpreted as quarterback passing **suppression**. **Alongside** the suppression of the quarterback in the passing game, a pattern in play comprising of defensive **disruptions were** established (TC7). Markers for this pattern of play manifested predominantly in sacks, tackles for loss and interceptions. Moreover, this feature comprised meaningful contributions from blitzes, hurries and quarterback knockdowns - **disruptive** defensive actions possibly leading to interception or turnover opportunities. Teams' capacity to contain ball carriers materialised in the final defensive factor (TC8). While Berri and Burke's (2011) appropriation of James' (1974) range factor has been proposed to predict and evaluate individual tackling efficacy (tackle factor), variability according to positional differences renders the metric impractical and limited in generalisability. Therefore, a pure measure of missed tackles associated with rushing and passing broken tackles in addition to defensive hurries provide a first depiction of team tackling performance in the NFL.

An understanding of composite performance indicator contributions has been hypothesised to elucidate the variability in observed strategic behaviours (Wedding et al., 2021). This exploration of playing styles in the National football league has demonstrated the presence of underlying structures in play based upon technical variables. More importantly, gaining an understanding of the factor loadings for variables contributions towards specific aspects of play enables analysts to weight values according to their relative importance, create composite scores, or team profiles to contextualize performance. As such, these styles of play may enable the navigation of evolving match dynamics and provide a means by which tactics may be manipulated to increase the likelihood of successful outcomes. Indeed, whereas previous research has determined the importance of total offensive and defensive prowess in the pass and run game (Cohea and Payton, 2011, May et al., 2010), this study has investigated

the characteristics of these patterns in play. Offensively, team efficacy on third down on pass, run and scrambling opportunities were associated with play types while demonstrating a clear importance within their respective styles of play. These findings contradict previous research by Cafarelli, Rigdon and Rigdon (2012), which identified third down success rates as almost meaningless. Notwithstanding, the ability to gain yards before and after contact as well as achieving yards after the catch provided meaningful additions to game styles. With the inclusion of uniquely defensive variables within exploratory factor analysis, teams appear to adopt three clear playing patterns wherein efforts are made towards the suppression **and defensive coverage** of the opposition passing game, containing ball-carriers in open play or penetrative efforts towards the quarterback - manifesting in **disruptions**. It has been established that mixed strategies are employed by NFL teams, confirming that game styles are not executed without regard for opponent actions (McGarrrity and Linnen, 2010). Therefore, further analysis aimed towards evaluating teams' ability to implement a variety of playing styles such as those identified in the present study may be used to quantify opposition strengths and weaknesses in specific areas of play. Moreover, in practice, the identified styles of play can be seen from both offensive and defensive perspectives. Hence, a combination of multivariate solutions, used to provide easily interpretable results may be used to inform strategic preparation for matchups, or adjust tactics during games.

4.1 Limitations and future research

With passing styles accounting for 48% of the proportional variance, scrambling and rushing factors 29%, and defensive features explaining 23%, greater detail and equal representations of performance concerning the variables included within analyses may result in a more elaborate explanation of playing styles in American football. The relative predictive power of the identified playing styles may therefore be restricted by the granularity of the data collected.

This effect is particularly notable in scrambling and defensive areas of play. Whereas Byanna and Klabjan (2016) produced spatiotemporal measures to quantify blocking performance, there is a dearth of publicly accessible technical metrics regarding offensive and defensive output. Furthermore, the complexity of these analyses remains challenging for practical implementation (Lord et al., 2020). This effect has been noted by Dutta et al. (2020), wherein a majority of data collected on NFL events is constrained to actions performed by, or towards the ball carrier - omitting essential information.

What's more, the identification of playing styles as understood in many sports may not apply directly to American football. Indeed, whereas certain principal components may be identified as a style of play, some may be considered aspects of play or latent areas of performance. Notwithstanding, while equivalent approaches are required to corroborate these findings in college football, the use of unsupervised techniques to discern the mechanisms underlying the variation in performance patterns provides a platform to understanding team interactions in professional American football. However, adopting these styles and the relative quality within them exhibited by teams may be harnessed to enhance evaluations of team performance in the NFL. Equally important, engaged alongside situational variables (e.g. match location, yard line, down) and game states, quality exhibited under several playing styles may inform the predictability of team and opposition behaviours with respect to realistic in-game scenarios (Gollan et al., 2020).

5 Conclusion

Whereas alternative football codes have engaged numerous methods to discern patterns in play and performance tendencies, no such investigations concerning American football have been published. This study empirically derived eight playing styles in the National Football League,

accounting for several areas of play through associations between technical variables. These predominantly uncorrelated latent variables provide a simple and interpretable structure from which team performance may be further analysed - providing a tangible tool for the design of specific training drills (Fernandez-Navarro et al., 2016), the accurate simulation of competition scenarios and the improvement of proactive offensive or defensive decision making by coaches (Sampaio et al., 2010).

References

- Amatria, M., Maneiro, R., Casal, C.A., Papadopoulou, S., Sarmiento, H., Ardá, A., Iglesias, X. and Losada, J.L., 2021. Differences in Technical Development and Playing Space in Three UEFA Champions Leagues. *Frontiers in psychology*, 12, p.695853.
- Bartlett, M.S., 1954. A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.296-298.
- Berri, D.J. and Burke, B., 2011. Measuring productivity of NFL players. In *The economics of the National Football League: The state of the art* (pp. 137-158). New York, NY: Springer New York.
- Biro, P. and Walker, S.G., 2022. A reinforcement learning based approach to play calling in football. *Journal of Quantitative Analysis in Sports*, 18(2), pp.97-112.
- Byanna, N. and Klabjan, D., 2016. Evaluating the Performance of Offensive Linemen in the NFL. *arXiv preprint arXiv:1603.07593*.
- Carter, V. and Machol, R.E., 1971. Operations research on football. *Operations Research*, 19(2), pp.541-544.
- Cattell, R.B., 1966. The scree test for the number of factors. *Multivariate behavioral research*, 1(2), pp.245-276.
- Cintia, P. and Pappalardo, L., 2021. Coach2vec: autoencoding the playing style of soccer coaches. *arXiv preprint arXiv:2106.15444*.
- Clijmans, J., Van Roy, M. and Davis, J., 2022, September. Looking Beyond the Past: Analyzing the Intrinsic Playing Style of Soccer Teams. In *Joint European Conference on*

Machine Learning and Knowledge Discovery in Databases (pp. 370-385). Cham: Springer Nature Switzerland.

Cohea, C. and Payton, M.E., 2011. Relationships between player actions and game outcomes in american football. *Sportscience*, 15, pp.19-25.

Comrey, A.L. and Lee, H.B., 2013. A first course in factor analysis. Psychology press.

Craig, J.D. and Winchester, N., 2021. Predicting the national football league potential of college quarterbacks. *European Journal of Operational Research*, 295(2), pp.733-743.

Decroos, T. and Davis, J., 2020. Player vectors: Characterizing soccer players' playing style from match event streams. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part III* (pp. 569-584). Springer International Publishing.

Deshpande, S.K. and Evans, K., 2020. Expected hypothetical completion probability. *Journal of Quantitative Analysis in Sports*, 16(2), pp.85-94.

Dutta, R., Yurko, R. and Ventura, S.L., 2020. Unsupervised methods for identifying pass coverage among defensive backs with NFL player tracking data. *Journal of Quantitative Analysis in Sports*, 16(2), pp.143-161.

Emara, N., Owens, D., Smith, J. and Wilmer, L., 2017. Serial correlation in National Football League play calling and its effects on outcomes. *Journal of behavioral and experimental economics*, 69, pp.125-132.

Fernandez-Navarro, J., Fradua, L., Zubillaga, A., Ford, P.R. and McRobert, A.P., 2016. Attacking and defensive styles of play in soccer: analysis of Spanish and English elite teams. *Journal of sports sciences*, 34(24), pp.2195-2204.

Fernandez-Navarro, J., Fradua, L., Zubillaga, A. and McRobert, A.P., 2018. Influence of contextual variables on styles of play in soccer. *International Journal of Performance Analysis in Sport*, 18(3), pp.423-436.

Field, A., 2013. *Discovering statistics using IBM SPSS statistics*. sage.

Fokoue, E. and Foehrenbach, D., 2013. A statistical data mining approach to determining the factors that distinguish championship caliber teams in the National Football League.

Gløersen, Ø., Myklebust, H., Hallén, J. and Federolf, P., 2018. Technique analysis in elite athletes using principal component analysis. *Journal of sports sciences*, 36(2), pp.229-237.

Gollan, S., Bellenger, C. and Norton, K., 2020. Contextual factors impact styles of play in the English premier league. *Journal of sports science & medicine*, 19(1), p.78.

Gollan, S., Ferrar, K. and Norton, K., 2018. Characterising game styles in the English Premier League using the “moments of play” framework. *International Journal of Performance Analysis in Sport*, 18(6), pp.998-1009.

Gómez, M.Á., Mitrotasios, M., Armatas, V. and Lago-Peñas, C., 2018. Analysis of playing styles according to team quality and match location in Greek professional soccer. *International Journal of Performance Analysis in Sport*, 18(6), pp.986-997.

Greenham, G., Hewitt, A. and Norton, K., 2017. A pilot study to measure game style within Australian football. *International Journal of Performance Analysis in Sport*, 17(4), pp.576-585.

Han, J., Kamber, M. and Pei, J., 2011. *Data mining concepts and techniques third edition*. The Morgan Kaufmann Series in Data Management Systems, 5(4), pp.83-124.

Haar, A.V., 2019. Exploratory data analysis of passing plays using nfl tracking data. NFL Big Data Bowl: Indianapolis, IN, USA.

Hewitt, A., 2007. Match analyses of Australian international women soccer players using an athlete tracking device. Performance analysis in soccer: applications of player tracking technology, p.65.

Ho, T., Carl, S., 2023. nflreadr: Download 'nflverse' Data. <https://nflreadr.nflverse.com>, <https://github.com/nflverse/nflreadr>.

Ibrahim, P., 2021. A Spatial Framework for Analyzing NFL Offensive Line Play.

James, B., 1977. Baseball Abstract: Featuring 18 Categories of Statistical Information That You Just Can't Find Anywhere Else. Lawrence, Kansas: Self-published.

Jeon, G. and Park, J., 2021. Characterizing patterns of scoring and ties in competitive sports. Physica A: Statistical Mechanics and its Applications, 565, p.125544.

Joash Fernandes, C., Yakubov, R., Li, Y., Prasad, A.K. and Chan, T.C., 2020. Predicting plays in the National Football League. Journal of Sports Analytics, 6(1), pp.35-43.

Jolliffe, I.T., 2002. Principal component analysis for special types of data (pp. 338-372). Springer New York.

Kong, L., Zhang, T., Zhou, C., Gomez, M.A., Hu, Y. and Zhang, S., 2022. The evaluation of playing styles integrating with contextual variables in professional soccer. Frontiers in Psychology, 13, p.1002566.

Lago-Peñas, C. and Dellal, A., 2010. Ball possession strategies in elite soccer according to the evolution of the match-score: the influence of situational variables. *Journal of human kinetics*, 25(2010), pp.93-100.

Lago-Peñas, C. and Gómez-López, M., 2014. How important is it to score a goal? The influence of the scoreline on match performance in elite soccer. *Perceptual and motor skills*, 119(3), pp.774-784.

Lago-Peñas, C., Gómez-Ruano, M. and Yang, G., 2017. Styles of play in professional soccer: an approach of the Chinese Soccer Super League. *International Journal of Performance Analysis in Sport*, 17(6), pp.1073-1084.

Lee, P., Chen, R. and Lakshman, V., 2016. Predicting offensive play types in the National Football League.

Lock, D. and Nettleton, D., 2014. Using random forests to estimate win probability before each play of an NFL game. *Journal of Quantitative Analysis in Sports*, 10(2), pp.197-205.

Lord, F., Pyne, D.B., Welvaert, M. and Mara, J.K., 2020. Methods of performance analysis in team invasion sports: A systematic review. *Journal of sports sciences*, 38(20), pp.2338-2349.

McGarrity, J.P. and Linnen, B., 2010. Pass or run: An empirical test of the matching pennies game using data from the National Football League. *Southern Economic Journal*, 76(3), pp.791-810.

McLean, S., Salmon, P.M., Gorman, A.D., Naughton, M. and Solomon, C., 2017. Do inter-continental playing styles exist? Using social network analysis to compare goals from the 2016 EURO and COPA football tournaments knock-out stages. *Theoretical Issues in Ergonomics Science*, 18(4), pp.370-383.

NFL (2021). URL: <https://www.nfl.com/scores/>

nflfastR (2021). URL: <https://github.com/guga31bb/nflfastR-data/tree/master/data>

Nguyen, Q., Yurko, R. and Matthews, G.J., 2023. Here Comes the STRAIN: Analyzing Defensive Pass Rush in American Football with Player Tracking Data. arXiv preprint arXiv:2305.10262.

Parmar, N., James, N., Hearne, G. and Jones, B., 2018. Using principal component analysis to develop performance indicators in professional rugby league. International Journal of Performance Analysis in Sport, 18(6), pp.938-949.

Plakias, S., Moustakidis, S., Kokkotis, C., Tsatalas, T., Papalexi, M., Plakias, D., Giakas, G. and Tsaopoulos, D., 2023. Identifying soccer teams' styles of play: a scoping and critical review. Journal of Functional Morphology and Kinesiology, 8(2), p.39.

Posit team (2023). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL: <http://www.posit.co/>.

Pro-football-reference, 2023. https://www.pro-football-reference.com/about/advanced_stats.htm

Raiche, G., Roipel, M. and Blais, J.G., 2006. Non-graphical solutions for the Catell's scree plot. In The International Annual Meeting of the Psychometric Society.

Raîche, G., Walls, T.A., Magis, D., Riopel, M. and Blais, J.G., 2013. Non-graphical solutions for Cattell's scree test. Methodology.

Ruan, L., Ge, H., Shen, Y., Pu, Z., Zong, S. and Cui, Y., 2022. Quantifying the effectiveness of defensive playing styles in the Chinese Football Super League. *Frontiers in Psychology*, p.2461.

Sampaio, J., Lago, C., Casais, L. and Leite, N., 2010. Effects of starting score-line, game location, and quality of opposition in basketball quarter score. *European Journal of Sport Science*, 10(6), pp.391-396.

Sarmiento, H., Marcelino, R., Anguera, M.T., Campaniço, J., Matos, N. and Leitão, J.C., 2014. Match analysis in football: a systematic review. *Journal of sports sciences*, 32(20), pp.1831-1843.

Sheehan, W.B., Tribolet, R., Watsford, M.L., Novak, A.R., Rennie, M. and Fransen, J., 2022. Tactical analysis of individual and team behaviour in professional Australian Football: Original investigation. *Science and Medicine in Football*, 6(2), pp.172-180.

Tabachnick, B.G. and Fidell, L.S., 2012. *Using Multivariate Statistics*.

Teich, B., Lutz, R. and Kassarnig, V., 2016. NFL play prediction. *arXiv preprint arXiv:1601.00574*.

Walsh, R., 2021. Identifying significant features for Player Evaluation in NFL comparing ANNs and Traditional Models.

Wedding, C., Woods, C.T., Sinclair, W.H., Gomez, M.A. and Leicht, A.S., 2021. Exploring the effect of various match factors on team playing styles in the National Rugby League. *International Journal of Sports Science & Coaching*, 16(4), pp.976-984.

Winter, C. and Pfeiffer, M., 2016. Tactical metrics that discriminate winning, drawing and losing teams in UEFA Euro 2012®. *Journal of sports sciences*, 34(6), pp.486-492.

Yurko, R., Ventura, S. and Horowitz, M., 2019. nflWAR: a reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports*, 15(3), pp.163-183.

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D. and Saeed, J., 2020. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, 1(2), pp.56-70.

Zuccolotto, P., Manisera, M. and Sandri, M., 2018. Big data analytics for modeling scoring probability in basketball: The effect of shooting under high-pressure conditions. *International journal of sports science & coaching*, 13(4), pp.569-589.

Figures

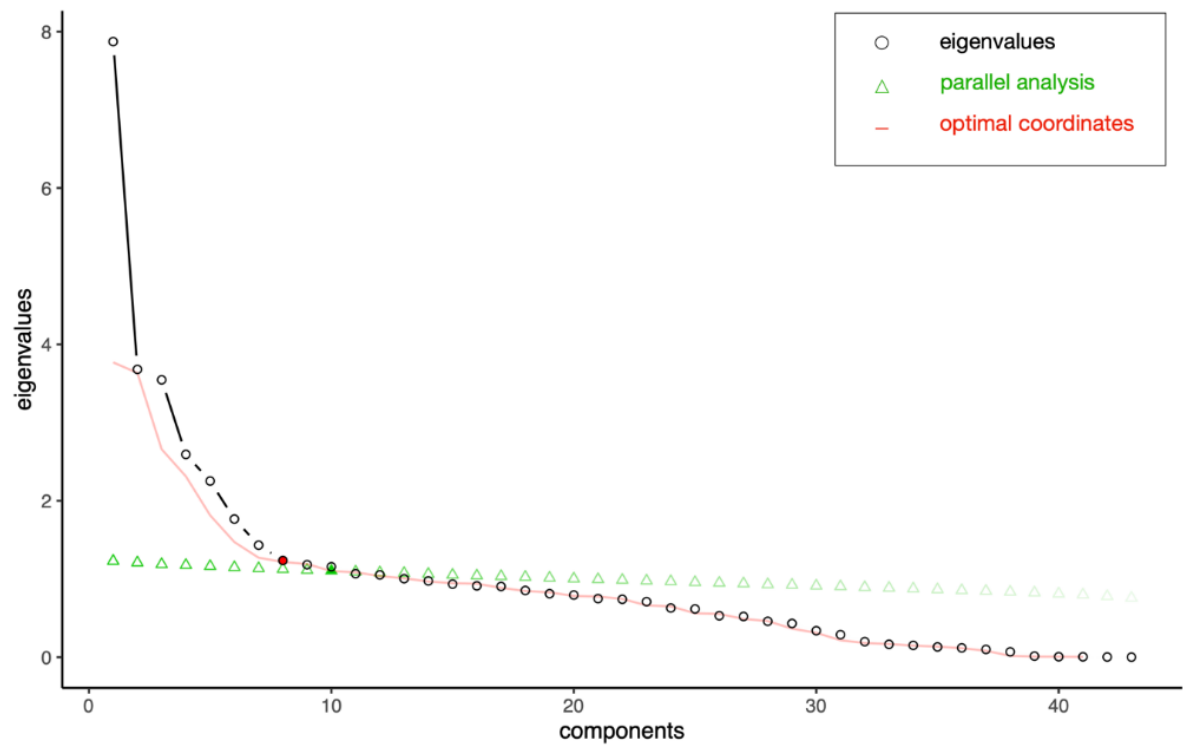


Figure 1. Scree test with non-graphical solutions (PCA component retention)

Tables

Table 1. Variance accounted for by principal components

Variance	TC2	TC6	TC3	TC4	TC1	TC5	TC7	TC8
Sum of square loadings	5.05	4.01	3.40	3.54	2.78	2.50	1.73	1.38
Proportional variance	0.12	0.09	0.08	0.08	0.06	0.06	0.04	0.03
Cumulative variance	0.12	0.21	0.29	0.37	0.44	0.49	0.53	0.57
Proportion explained	0.21	0.16	0.14	0.15	0.11	0.10	0.07	0.06
Cumulative proportion	0.21	0.37	0.51	0.66	0.77	0.87	0.94	1.00

Table 2. Factor loadings of PCA with an oblique, Direct Oblimin rotation.

Variables	TC2	TC6	TC3	TC4	TC1	TC5	TC7	TC8
completed passes	0.90							
passing first downs	0.83							
passing yards after catch	0.82							
short passes	0.77				0.39			
passing yards	0.76							
pass attempts	0.61				0.38			
quarterback drops	0.61				0.38			
successful passing third downs	0.48							
receiving tackles broken	0.46							0.41
deep passes		0.97						
intended passing air yards		0.92						
passing air yards completed		0.92						
inaccurate passes		0.58			0.42			
scramble yards			0.92					
scramble attempts			0.92					
scrambling first downs			0.91					
successful scrambling third downs			0.72					
unsuccessful scrambling third downs			0.38					
rushing yards				0.94				
rushing first downs				0.89				
rushing attempts				0.79				
rushing yards before contact				0.74				
rushing yards after contact				0.71				
successful rushing third downs				0.61				
unsuccessful rushing third downs				0.50				
unsuccessful passing third downs					0.81			
incomplete passes		0.54			0.58			
quarterback hits					0.49			
unsuccessful passing fourth downs					0.39			
receiving dropped passes					0.35			
defensive completed air yards allowed						0.66		
defensive yards after catch allowed						0.64		
defensive quarterback knockdowns						0.49	0.39	
defensive combined tackles	-0.37			-0.41		0.47		
defensive blitzes						0.39	0.36	
defensive sacks							0.61	
defensive tackles for loss							0.51	
defensive interceptions							0.47	
rushing tackles broken				0.38				0.60
defensive missed tackles								0.58
defensive hurries							0.36	0.42

Principal components ordered according to proportional variance explained. Factor loadings $<|0.32|$ were rejected as they account for less than approximately 10% of the variance within a component. Variables with no association to styles of play: successful passing fourth downs, successful scrambling fourth downs, successful rushing fourth down, unsuccessful scrambling fourth downs, unsuccessful rushing fourth down.

Table 3. Principal components correlation matrix.

PC	TC2	TC6	TC3	TC4	TC1	TC5	TC7	TC8
TC2		0.35	0.02	-0.16	0.12	-0.07	-0.05	-0.04
TC6			0.06	-0.25	0.25	0.08	-0.11	-0.03
TC3				0.08	0.09	-0.03	-0.08	-0.04
TC4					-0.21	-0.12	0.19	0.04
TC1						-0.11	-0.11	-0.05
TC5							0.00	0.07
TC7								0.04
TC8								

Where TC2 = passing efficacy; TC6 = deep passing; TC3 = scrambling; TC4 = rushing; TC1 = passing efficiency; TC5 = quarterback suppression; TC7 = defensive disruption and TC8 = tackling.

A Performance variable operational definitions

Variable	Operational definition
Complete pass	Successful pass attempts.
Deep passes	Total pass attempts achieving 15 or more air yards.
Defensive blitzes	Frequency of defensive plays where 5+ players or a defensive back attempted to reach the quarterback.
Defensive combined tackles	The combination of solo and assisted tackles completed by the defence.
Defensive completed air yards allowed	Total completed passing air yards conceded by the defence.
Defensive hurries	Frequency of passing attempted earlier than intended or the quarterback was chased out of the pocket.
Defensive missed tackles	The combination of solo and assisted tackles failed by the defence.
Defensive quarterback knockdowns	Frequency that defenders make contact with the quarterback - forcing them to the ground after a pass attempt.
Defensive sacks	Frequency of tackles made on the passer or quarterback whilst in possession of the ball.
Defensive Tackles for loss	Frequency of ball carriers tackled behind the line of scrimmage on pass, scramble or rush attempts.
Defensive yards after catch allowed	Total yards after the catch conceded by the defence.
Inaccurate passes	Total pass attempts that were not catchable by the receiver with 'normal' effort.
Incomplete passes	Total attempts to pass the ball forwards that did not result in a catch.
Intended passing air yards	Total distance achieved in the air on passing attempts prior to an attempted catch or pass incomplection.
Pass attempts	Total attempts to pass the ball forwards.
Passing air yards completed	Total distance achieved in the air on successful catches.
Passing first downs	Frequency of first downs achieved on passing attempts.
Passing yards	The combination of air yards achieved and yards after the catch achieved on completed passing attempts.
Passing yards after catch	Total yards achieved by a receiver after catching a pass attempt.
Quarterback drop	Frequency of plays where the ball was snapped to the quarterback directly behind (in close proximity to) the centre and dropping back with the intent to attempt a pass.
Quarterback hits	Total contacts made defensive players on the quarterback at any time during the play.
Quarterback scramble	Intended passing play resulting in an improvised run by the quarterback reaching beyond the line of scrimmage.
Receiving dropped passes	Total opportunities to catch the ball from passing attempts dropped by the receiver.
Receiving tackles broken	Total tackles broken during pass receiver ball carries.
Rush attempts	Designed running play where no forward pass was attempted.
Rushing first downs	Frequency of first downs achieved on designed run attempts.
Rushing tackles broken	Total tackles broken during rush attempts by ball carriers.
Rushing yards	Total yards achieved during rushing attempts.
Rushing yards after contact	Total yards achieved during rushing attempts after any defensive player makes contact with the ball carrier.
Rushing yards before contact	Total yards achieved during rushing attempts before any defensive player makes contact with the ball carrier.
Scramble yards	Total yards achieved from scramble attempts

Scrambling first downs	Frequency of first downs achieved on scrambling attempts.
Short passes	Total pass attempts achieving 15 or fewer air yards.
Successful passing fourth downs	Total fourth down opportunities converted from passing attempts.
Successful passing third downs	Total third down opportunities converted from passing attempts.
Successful rushing fourth downs	Total fourth down opportunities converted from designed run attempts.
Successful rushing third downs	Total third down opportunities converted from designed run attempts.
Successful scrambling fourth downs	Total fourth down opportunities converted from scrambling attempts.
Successful scrambling third downs	Frequency of third downs converted by scrambling attempts.
Unsuccessful passing fourth downs	Total fourth down opportunities failed from passing attempts.
Unsuccessful passing third downs	Total third down opportunities failed from passing attempts.
Unsuccessful rushing fourth downs	Total fourth down opportunities failed from designed run attempts.
Unsuccessful rushing third downs	Total third down opportunities failed from designed run attempts.
Unsuccessful scrambling fourth downs	Total fourth down opportunities failed from scrambling attempts.
Unsuccessful scrambling third downs	Total third down opportunities failed from scrambling attempts.

B Ethical approval letter



London Sport Institute REC

The Burroughs
Hendon
London NW4 4BT

Main Switchboard: 0208 411 5000

24/11/2021

APPLICATION NUMBER: 18704

Dear Ben Adam Carter and all collaborators/co-investigators

Re your application title: Play style proficiency on play calling tendencies (NFL).

Supervisor:

Co-investigators/collaborators:

Thank you for submitting your application. I can confirm that your application has been given APPROVAL from the date of this letter by the London Sport Institute REC.

The following documents have been reviewed and approved as part of this research ethics application:

Document Type	File Name	Date	Version
Aims, objectives and hypotheses	Ben Carter - The influence of play style proficiency on play calling tendencies in the National Football League (NFL). - Project Outline Template	14/09/2021	1
Methods and data	Ben Carter - The influence of play style proficiency on play calling tendencies in the National Football League (NFL). - Research Procedure Template	14/09/2021	1

Although your application has been approved, the reviewers of your application may have made some useful comments on your application. Please look at your online application again to check whether the reviewers have added any comments for you to look at.

Also, please note the following:

1. Please ensure that you contact your supervisor/research ethics committee (REC) if any changes are made to the research project which could affect your ethics approval. There is an Amendment sub-form on MORE that can be completed and submitted to your REC for further review.
2. You must notify your supervisor/REC if there is a breach in data protection management or any issues that arise that may lead to a health and safety concern or conflict of interests.
3. If you require more time to complete your research, i.e., beyond the date specified in your application, please complete the Extension sub-form on MORE and submit it your REC for review.
4. Please quote the application number in any correspondence.
5. It is important that you retain this document as evidence of research ethics approval, as it may be required for submission to external bodies (e.g., NHS, grant awarding bodies) or as part of your research report, dissemination (e.g., journal articles) and data management plan.
6. Also, please forward any other information that would be helpful in enhancing our application form and procedures - please contact MOREsupport@mdx.ac.uk to provide feedback.

Good luck with your research.

Yours sincerely,

Chairs Dr Rhonda Cohen/ Dr Anne Elliott

London Sport Institute REC