Book Section:
Gerrard, B (2017) Analytics, Technology and High Performance Sport. In: Schulenkorf, N and Frawley, S, (eds.) Critical Issues in Global Sport Management. Routledge , London; New York . ISBN 9781138911222

Chapter 14: Analytics, Technology and High Performance Sport

Bill Gerrard

Leeds University Business School, UK

June 2015; Revised November 2015

To be published in:

N. Schulenkorf and S. Frawley (eds),

Critical Issues in Global Sport Management

(London: Routledge)

Abstract

This chapter considers how technological developments have impacted on the use of data to manage performance in sport. In particular the move from paper-and-pencil methods of data collection to video recording and then, more recently, to the use of digital technology has both increased the range and detail of the data collected. This has resulted in the emergence of expert data systems specific to individual teams based on their own tactical game plans with coaches providing systematic evaluations of the decision making and technique of their players. The greater availability of performance data has seen teams start to use statistical analysis to support decision making. The two principal functions of statistical analysis are distinguished – the reporting function and the analytical function – and it is argued that the reporting function is dependent on the analytical function to identify which metrics should be reported and tracked.. The increasing use of sports analytics is illustrated by the examples of the Oakland A's in baseball, Bolton Wanderers in association football, and Saracens in rugby union.

Keywords: expert data systems; key performance indicators; sabermetrics; Moneyball; invasion-territorial team sports

Chapter 14: Analytics, Technology and High Performance Sport

Performance analysis is at the core of the coaching process in sport at all levels of performance. Coaches attempt to facilitate improvements in the future performance of their athletes by investigating the key factors involved in the outcomes of recent games. This performance analysis involves analysing the observed outcomes of performances to understand the underlying causal mechanisms. Understanding the deeper structure of performances, particularly identifying the controllables—the key factors that are under the control of the athlete—is crucial in order for the coach to decide on the optimal intervention strategy to facilitate improved future performance. For example, what should be the training priorities in preparation for the next contest? What is the best tactical plan for the next contest? In the case of team sports, who should be selected to participate in the next contest? These are all critical decisions for a coach that may have a significant impact on the outcome of the next contest in which their athletes compete, and all of these coaching decisions can be informed by performance analysis.

This chapter explores how technological developments have revolutionised the collection of performance data with profound consequences for performance analysis. It is shown how the advent of video recording and then digital technology has had two effects on the type of performance data collected. Firstly, there has been a widening of the range of observed actions on which data can be collected. Secondly, there has also been a deepening of the nature of the data collected with technological advances providing the means to capture data on more fundamental aspects of performance. Performance analysis not only includes frequencies of different types of actions that have been performed, but also data on the more fundamental dimensions of performance underlying the observed actions. These more fundamental aspects of performance can be summarised as the four A's: ability (i.e., technical), athleticism (i.e., physical), attitude (i.e., psychological) and awareness (i.e., tactical). Recent technological developments have particularly impacted the collection of tracking data on distances covered and speed of an athlete's movement. The impact of technology on performance data collection is discussed in this chapter using a three-stage schema: (a) paper-and-pencil methods, (b) video recording and (c) digital technology. The enhanced possibilities for greater use of expert data to assess the more fundamental aspects of performance are also considered.

The impact of technology on data collection in high performance sport has not only led to an exponential growth in the quantity of data collected, but has also massively increased the scope for statistical analysis of performance data. There has been a parallel development in the type of data analysis—qualitative to quantitative—used in high performance sport. Qualitative analysis based on video replays and the reconstruction of critical incidents has been supplemented by quantitative analysis—initially the reporting of summary performance statistics and, more recently, the use of statistical analysis and other related quantitative techniques—to analyse patterns across performances. Sport analytics depends on the use of statistical and other quantitative analytical techniques to inform the decision-making of elite coaches. The growth in sports analytics is surveyed in this chapter with particular emphasis on the distinction between statistical reporting and analytics.

Throughout the chapter the focus is on the specific context of team sports; however, we argue that the managerial implications are transferable to all sports and, indeed, to the management of non-sporting performance. Of particular importance is the distinction between the striking and fielding team sports, such as cricket and baseball, and the invasion-territorial team sports, such as the various codes of football. Further, we argue that the technological developments in data collection, for studies on performance, have impacted most significantly the invasion-territorial sports with the widespread use of count data in these sports only really possible with the advent of digital technology. We conclude the chapter with a summary of the key points and some thoughts on the outlook for performance analysis and spots analytics.

The Impact of Technology on Performance Analysis

A useful way of understanding the impact in recent years of technological developments on the collection and analysis of performance data is to consider three broad stages starting with the use of paper-and-pencil methods and then progressing to the impact of video recording and digital technology. A key aspect in this evolution of performance analysis is that the impact of video recording and digital technology has been much greater in the invasion-territorial sports, such as football.

Stage 1: Paper-and-Pencil Methods

Striking and fielding sports—primarily cricket and baseball—are structurally very simple games; at the core of cricket and baseball is a sequence of head-to-head contests between the batsman/hitter and the bowler/pitcher. The structure of striking and fielding sports has two very important implications for performance analysis; for example, the performance of athletes in striking and fielding sports is highly separable such that the major contributions of individual players can be easily identified and separated from the contributions of other players. Further, there is a very direct link between the contributions of individual players and match outcomes. Indeed, match outcomes are largely driven by the performances of the batsman/hitter and the bowler/pitcher. The defensive contributions of fielders in preventing runs and getting players out through catching or returning the ball to the wicket/base ahead of a running player are relatively minor in comparison to the contributions of the bowler/pitcher;

however, the interdependence of bowling/pitching and fielding remains an issue in evaluating defensive contributions particularly in baseball.

The structure of striking and fielding sports as a sequence of individual contests has facilitated the collection of key performance data during games; play is naturally broken up between bowls/pitches as the ball is returned to the bowler/pitcher in preparation for the next delivery giving time for the basic details of each play to be recorded on a scorecard. These scorecards are of a “double-entry” form, with each play recorded from the perspectives of both the batsman/hitter and the bowler/pitcher. In cricket, the batting team’s scorecard shows the outcome for each batsman for every ball faced as well as how the batsman’s innings ended while the bowling team’s scorecard shows the outcome of every delivery by each of their bowlers. The natural delay between deliveries indicates sufficient time to record these outcomes. From the scorecards basic count data of runs scored by each batsman, wickets taken and runs conceded by each bowler can be calculated easily and these, in turn, can be used to generate batting averages (i.e., runs scored per innings) and bowling averages (i.e., runs conceded per wicket taken) for comparisons across games. Cross-checking for internal consistency between the batting and bowling scorecards and the match score should ensure that the data collected is accurate, free from recording errors. The scorecards should provide an objective record of a match since the data reports the outcomes as adjudicated and signalled by the umpire. A similar scoring system was developed for baseball with the format of box scores being originally adapted from the cricketing scorecard by Henry Chadwick in the 1860s, albeit allowing for the additional complexities in baseball for how runs are scored, particularly runs batted in, as well as whether or not a pitch is judged by the umpire as a strike or a ball, given that three strikes is an out whereas four balls advances the hitter to first base on a walk (Schwartz, 2004).

Stage 2: Video Recording

Paper-and-pencil methods, although highly effective for data collection in striking and fielding sports, are of much more limited use in invasion-territorial sports. These are the team sports, such as the various codes of football, hockey and basketball, that attempt to replicate the battlefield with teams scoring by moving an object (e.g., a ball or puck) to designated locations (e.g., between posts, across a line or through a hoop) defended by the opposing team. In addition to being territorial, these games are also time-dependent with the match outcome determined by the cumulative scores within a specified playing time. Invasion-territorial sports by their very nature are highly tactical; the game cannot be reduced easily to a sequence of head-to-head contests between individual players, as in striking and fielding sports. All players on the field are potentially involved in every play and, to be effective, all players on a team must coordinate their actions. Therefore, players need to be tactically and spatially aware in their decision-making during games and this becomes a key aspect in the analysis of player performance (see Section 3 below).

Invasion-territorial sports are much more complex in structure than striking and fielding games, which creates greater challenges for both data collection and data analysis. There is a much more indirect link between the individual contributions of players, scoring and match outcomes. Attacking and defending are much more interdependent, since possession of the ball, or puck, by the attacking team also has a defensive function by denying attacking opportunities to the opposing team as well as running down the clock. Possession and time are the ultimate scarce resources in invasion-territorial games. Attacking and defending are also interdependent in the sense that in most invasion-territorial games, possession has to be won by the defending team before an attacking play can instigate. Within attacking plays, different actions are also highly interdependent; in order to score, the ball must be moved forward into areas of the pitch in which scoring opportunities can be created, and these opportunities must in turn be converted into scores. This structural hierarchy of interdependent player actions in invasion-territorial sports creates a real separability issue for the performance analyst of how to weight the importance of different individual contributions and becomes even more complicated when certain contributions involve joint actions, such as lineouts and scrums in rugby union.

The nature of invasion-territorial sports complicates the data collection process. For example, in striking and fielding sports, games need to be coded principally in regard to two dimensions—players (Who?) and actions (What?)—while in invasion-territorial games, there are four basic dimensions: players (Who?), actions (What?), position (Where?) and time (When?). But not only is the data collection process made more demanding by the need to gather a greater amount of information both in terms of the number of dimensions and the variety of possible alternatives within each dimension, there is also much greater time pressure. Unlike striking and fielding sports, there is often little or no natural delay between individual plays, with teams switching continuously between attacking and defending as possession changes.

Not surprisingly, paper-and-pencil methods can only capture very limited amounts of performance data during invasion-territorial games. Historically, the only data consistently recorded in invasion-territorial sports have been appearances, scores, disciplinary warnings and player ejections (e.g., sin-bins and sending-offs). The advent of the video recorder in the 1970s—with the ability to pause, rewind and fast forward game videos—greatly facilitated the ability to collect performance data in invasion-territorial sports. Although coaches had, for many years, watched films of games to review performances and analyse future opponents, the use of game films had been restricted by the limited availability of game films, the need for projection facilities to watch these games and the difficulties in replaying specific critical moments. The explosion in the TV coverage of sport from the 1960s onwards combined with the development of video recording technology in the 1970s resolved all of these issues and opened up the opportunity for detailed video analysis of games. Rather than being restricted to in-game data collection, teams could now compile performance data by applying paper-and-pencil methods to video replays of the games. The process was slow and time-consuming, so, as a consequence, teams tended to focus only on critical incidents. Thus, video analysis tended to be largely qualitative with reconstructions of specific plays, often those in which a team had scored or had conceded a score. However, teams identified patterns of play based on very few observations, indicating a sample size limitation. Count data for invasion-territorial sports remained very limited in scope.

Stage 3: Digital Technology

The advent of digital technology and low-cost computing power in the 1990s accelerated the process of data collection in high performance sport. The digitalisation of video images allowed the development of computer software to code and edit game videos so that the data collection process became much quicker. In particular, the use of image-recognition software indicated that the coding of games could be largely automated, with human intervention required only to resolve problems such as blind spots and conflicts when two or more players converged. Because of digital technology, count data for all of the different player actions in invasion-territorial sports became readily available.

The digital revolution in performance analysis led a number of commercial companies to offer data collection and analysis to elite sports. For example, in association football in the United Kingdom, three alternative commercial data systems have been available since the late 1990s: Opta, Prozone and Amisco. All these data systems provide detailed count data on all aspects of games. Opta also supplies the detailed coding of games in spreadsheets involving around 1,600 rows of data for the typical games. Further, Prozone and Amisco combine the provision of the count data with graphic representations and the game video. Both of these systems also allow some degree of interactivity such that the analyst is able to control some of the parameters for reporting the data. Typically, teams rely on these systems to provide detailed statistical reports after the game, for being a source of video and count data on future opponents and possible scouting targets for player recruitment. Teams tend to produce their own in-game data using systems such as SportsCode to code and edit the live video stream of their games. Given the obvious time pressures, analysts can code relatively few features of a game in real time and, therefore, focus only on data of particular relevance to coaches and mostly likely to influence tactical changes. Sam Allardyce was an early adopter of in-game data collection and video analysis in association football in England when he was manager of Bolton Wanderers between 1999 and 2007. At home games in the Reebok Stadium, Bolton's performance analysts collected performance data and edited the video during games. A Smartboard was installed in the home dressing room so that players could be shown video clips during the halftime interval with a particular emphasis on set-piece play, a key aspect of Bolton's game plan. Allardyce would watch most of the game from a seat in the stands rather than from the dugout at pitchside in order to get a better view of the game and would relay instructions to the analysts and coaches through an intercom system.

The development of digital performance systems not only led to a massive expansion in the provision of count data in invasion-territorial sports, but also to the availability of tracking data. Tracking data involves continuous data on the position of players on the pitch. Initially, this type of data was collected from video-based image recognition systems using triangulation of multiple cameras to continually locate the exact position of all players. This allowed distances covered and, when combined with time data, speeds to be calculated. For the first time, teams had reliable data on one aspect of the physical performance of players during games.

An alternative approach to video-based tracking systems is to use global positioning systems (GPS). Australian Rules Football was a first mover in this respect, allowing four players on either team to wear GPS devices during games in the 2005 games. Other sports, such as rugby union, have adopted GPS devices to provide in-game tracking data. However, possibly the most innovative use of GPS devices is to track distances covered and speeds in training sessions.

Expert Data Systems

The advances in data collection methods in high-performance sport, particularly the availability of software such as SportsCode, allows teams themselves to code and edit game videos quickly and has created the possibility of expert data systems. Expert data systems refer to systems in which data are defined and interpreted by coaches rather than relying on third-party commercial providers. The following example from association football can illustrate the difference between count data and expert data: Suppose a player (A1) has possession of the ball and plays it long beyond one of his team mates (A2) positioned out wide who is marked by a defending player (D1). D1 turns to chase the ball but is closed down quickly by A2 and forced to put the ball into touch. The attacking team now has a throw-in deep in the opposition half near the penalty area. How would this be coded? Typically, count data produced by third-party commercial providers would code the long ball by A1 as an incomplete pass, to be regarded as a negative since it led to a loss of possession. Further, D1 would most likely be credited with a clearance. A2's contribution would often not be coded at all unless it involved an attempted tackle and, again, since A2 did not win possession directly, the tackle would be designated as a missed tackle and treated as a negative. Therefore, the play could be quite easily coded as two negatives for the attacking team (i.e., an incomplete pass and a missed tackle) and a positive for the defending team (i.e., a clearance). However, the attacking team has gained "second-ball" possession (i.e., a throw-in) in a dangerous area of the pitch close to the opposition penalty area. Indeed, this was a tactic frequently employed by Bolton Wanderers in the English Premiership under Sam Allardyce. If this incident was coded internally by the Bolton analysts based on expert knowledge of the coaches' tactical plan, then both A1 and A2 would have been credited with positive contributions—A1 for the well-placed long ball and A2 for closing down D1—forcing a thrown-in to be conceded. This example highlights that count data is far from being an objective reporting of frequencies of different actions. The definitions of actions involve interpretation and experts with more detailed tactical and technical knowledge may have a very different interpretation of an observed event than an external analyst applying a standardised set of generic definitions.

The example of second-ball possession also raises another key aspect of expert data systems: Expert data can go beyond the observed event to analyse and interpret the more fundamental dimensions of performance. In particular, both the long ball played by A1 and the chase by A2 involve tactical decisions and their technical execution. A1's decision to play the long ball needs to be evaluated relative to the other available options. Further, the technical execution of the long ball in terms of distance and speed needs to be assessed relative to the positioning of both A2 and D1. If the ball is hit too long and/or too fast, there would be no opportunity for an effective chase by A2. Similarly, A2's decision to chase and the technical effectiveness of that chase in forcing D1 to put the ball into touch can also be expertly evaluated. Coaches can assess both the selection of actions (i.e., the decision) and their execution (i.e., the technique). Ultimately, it is all about "doing the right things" (i.e., selection) and "doing things right" (i.e., execution). Only the team's experts—the coaches and the analysts—can evaluate the tactical and technical aspects of their own players' performances.

An important component of tactical decision-making in invasion-territorial sports is the spatial positioning of players. Sports players continuously have to make decisions on where to position themselves both when their team is attacking and defending. Attacking players need to provide options for the player in possession either by moving into positions to receive the ball or by making decoy runs to create space. Defenders must make positioning decisions in order to cover either a specific attacker or a specific area of the pitch. Tracking technology has greatly aided the ability of coaches to assess the positioning decisions of players. For example, Prozone provides an animated reconstruction of a game with a bird's eye view of the positions of players on the pitch, using a video-based tracking system. Connecting lines between players can be added to the animation to facilitate a better understanding of how players are coordinating their movements. This feature is particularly useful in association football to check the alignment of the back line of defenders. This feature also helped a leading English Premiership team to recognise that one of their defenders, an established international player, often lagged behind the rest of the back line when they moved out quickly and, therefore, undermined any attempt to catch opponents offside. Once the player was made aware of the problem using video clips and game animations, he rectified this aspect of his game and started to take greater responsibility for calling this defensive play.

One of the teams that embraced expert data systems was Saracens—one of the leading rugby union teams in England. Under the leadership of their Director of Rugby, Brendan Venter, a former South African international and a qualified medical practitioner well-versed in evidence-based approaches to decision-making, Saracens adopted a coach-led system of evidence-based performance management in 2009. Each coach has specific areas of responsibility and, after every game, collates the data on these aspects of player performance using internal definitions based on the team's own tactical approach. The data include assessments of key decisions by individual players, as well as the technical execution of particular actions. From this internal data, a set of key-performance indicators (KPIs) for the team as a whole as well as for individual players has been identified and these are monitored across the season using a traffic-lights system classifying the KPIs as excellent (green), satisfactory (amber) or poor (red). The KPIs are used to evaluate team and player performances and feed into coaching decisions on training priorities, game tactics and team selection. In contrast, when it comes to opposition analysis, Saracens relies on count data supplied by Opta for the quantitative data analysis of future opponents. Time and cost constraints make it infeasible to collate expert data on opponents beyond the detailed qualitative analysis of the videos of their most recent games.

The Development of Sports Analytics

Sports analytics is the use of quantitative data analysis of performance data to support coaching decisions. Sports analytics is not only the analysis of performance data, but also analysis with an explicit practical purpose. Specifically, sports analytics is analysis directed towards informing coaches' decisions on the optimal choice of intervention regarding, for example, training programmes, game preparation and player recruitment.

There has been a long history of the statistical analysis of performance data in sport—particularly in baseball where paper-and-pencil methods have captured most of the key data from the very early days of the professional game in the 1860s. Important landmarks in the analysis of baseball include the contributions of Evers and Fullerton (1910), Rickey (1954), Cook (1964) and Mills and Mills (1970). Further, Scully (1974) used baseball data to show how the economic value of player performance could be calculated. For example, Scully developed a two-stage procedure in which he first estimated the win contribution of players by applying regression analysis to quantify the relationship between player performance and the team win percentage using the slugging average to measure hitter performance and the strikeout-to-walk ratio to measure pitcher performance. Using regression analysis, Scully then estimated relationship between team win percentage and team revenues to calculate the financial value of an incremental change in team performance and, by implication, the financial value of a player's contribution to the team win percentage.

Probably the most influential contributor to the development of the statistical analysis of baseball—or "sabermetrics" based on the acronym for the Society for American Baseball Research—is Bill James, who first started publishing articles on baseball in the mid-1970s and launched his Baseball Abstract in 1977. James questioned the usefulness of conventional performance metrics for hitting, pitching and fielding and provided a number of important advances. One of his most important contributions has been including walks in the evaluation of hitting performance. James showed that on-base percentage, the percentage of at-bats in which a hitter gets to base, is a much better predictor of team success than either batting or slugging averages which only considered hits. Traditionally, walks had been perceived as a pitcher error but the ability of hitters to select which pitches to leave is a skill in itself, yet one ignored by the conventional hitting metrics and, as a consequence, largely ignored by teams.

None of these contributions to sabermetrics constituted analytics at the time since they were not motivated by any intention to influence decision makers in the game. The following classified advertisement in The Sporting News for the first Baseball Abstract by Bill James in 1977 gives a clear sense of the somewhat whimsical nature of the enterprise: “The 1977 Baseball Abstract: Contains 18 Statistical Categories That You Just Can’t Find Anywhere Else, And a New Table Baseball Game.” However, the Baseball Abstract led to the development of sports analytics when MLB teams started to take account of the insights of James and other sabermetricians. The team that led the way in using sabermetrics was the Oakland Athletics, who had pioneered the provision of statistics to fans in the early 1980s with STATS Inc. (Schwartz, 2004). The Oakland A’s—a small-market MLB team with a very restricted player budget—then took the lead in the application of sabermetrics to player recruitment in the mid-1990s, initially under the General Manager, Sandy Alderson, and then most famously, by his assistant General Manager and successor, Billy Beane. The story of how the A’s used analytics as a “David” strategy to compete with resource-richer rivals, such as the New York Yankees, is recounted in Michael Lewis’s (2003) book Moneyball: The Art of Winning an Unfair Game, which was subsequently turned into a Hollywood movie with Brad Pitt starring as Billy Beane.

Moneyball was a game-changer in high-performance sport, with an influence stretching well beyond baseball. The timing was incredibly fortuitous since the publication of Moneyball coincided with the beginning of the widespread use of digital technology for data collection and the establishment of commercial companies providing performance data in the invasion-territorial sports particularly association football and the other North American major leagues, including American football, ice hockey and basketball. Nothing persuades more in sport than winning so inevitability the success of the A's led coaches and analysts in other team sports to ask whether or not the Moneyball approach was transferable particularly to the invasion-territorial sports (Gerrard, 2007). Proponents of a statistical approach to performance analysis viewed the complexity of invasion-territorial sports as an opportunity rather than a barrier. From a statistical perspective, the high degree of interdependency coupled with a multitude of different types of player actions is ideally suited to multivariate analysis, such as multiple regression, to separate out the effects of individual actions.

One of the main functions of statistics is the reporting function—the purely descriptive task of summarising data using measures of central tendency (e.g., mean and median), dispersion (e.g., standard deviation and interquartile range) and distributional shape (e.g., skewness and kurtosis). The reporting function should also include the provision of comparisons between different metrics as well as comparisons of the same metrics for different entities and/or different time periods. The other principal function of statistics is the analytical function which involves the analysis of the observed variation in the data to determine if the variation is systematic or random, and then proceeding to investigate the factors associated with the systematic variation. These two functions are interlinked. Statistical analysis of the data, particularly the identification of the most significant factors associated with the systematic variation in match outcomes, is necessary in order to determine which KPIs should be reported and tracked. Statistical reporting of summary statistics and comparisons for selected performance metrics does not constitute analytics. What to report becomes the critical issue especially in an era where technology allows the capture of data on all aspects of performance and databases can be programmed to automatically generate statistical reports. Coaches increasingly face the problem of information overload with so much data available and thus the primary role of the statistical analyst or data scientist is to determine the most appropriate data for the specific purpose. Statistical reports on performance must always include reliable statistical analysis that combines rigour and relevance.

The provision of tracking data has opened up new challenges for sports analytics. One line of enquiry is the link between player and team performance as measured by count data on player contributions and the physical data on speed and distance covered. In much the same way as in association football and many other invasion-territorial sports, there is often a weak relationship between the share of possession and match outcomes, and there is no unambiguous link between distances covered—and speed of movement—and match outcomes. Many teams track high-intensity distance (i.e., distances covered at sprint speeds) as a KPI, but there is no necessary link between high-intensity distance and effective performance. For example, players with excellent tactical awareness may be quicker and more effective in their positional decision-making and thereby reduce the need to sprint to get into the optimal position. In the case of distance covered and speed of movement, it is often the same principle as with the share of possession: It is not the quantity that counts as much as the quality of the physical work rate and player contributions. Additionally, quality can only really be captured by expert data not count data.

A second challenge in tracking data is determining how best to summarise and report the available data on spatial positioning of players. As already noted, this data can allow coaches to assess the effectiveness of the positioning decisions of players; therefore, it is important to find the best ways to present this data to provide a reliable representation of the spatial distribution of players. One of the sports in which this work has progressed most is basketball. Shea (2014) suggested two key metrics for summarising the spatial distribution of basketball players: (a) the convex hull area, which measures the perimeter of the five-sided polygon created by the players in an a team; and (b) the distance from average, which measures the average distance of the players from the central point of the polygon. Shea defined these measures for both the attacking team and the defending team and then analyses the interaction. One particularly innovative application is to analyse how individual attacking players affect the spatial distribution of the defensive team—which Shea calls "defensive stretch." Shea also provided an example in which two attacking players, Allen and Wade, are compared. Although Wade averaged 6 inches further away than Allen from the attacking team's central point, it was in fact Allen who had the bigger impact on the defensive stretch of the opponents since Allen's defender averaged 14 inches more from the centre of the defending team than Wade's defender.

Managerial Implications

The overall message of this chapter is that the technology now exists to collect comprehensive and very detailed data on all aspects of athlete performance even in the most complex sports such as the invasion-territorial sports. "Knowing the numbers" is now a critical component in an effective competitive strategy for organisations operating in any sector, not just high-performance sport (Davenport & Harris, 2007). It is the growing recognition of the strategic importance of developing a leading-edge capability in analytics that accounts for the universal appeal of the Moneyball story of how a small resource-constrained organisation managed to compete effectively by knowing the numbers and using them better than their competitors. The Oakland Athletics provide an example for any organisation that wants to compete effectively in an ever-changing environment particularly when confronting resource-rich rivals.

For analytics to become an effective competitive tool, an organisation must develop a culture that values an evidence-based approach to decision-making and is, therefore, supportive of the use of data analysis. This culture requires leadership and senior management that is fully committed to building an analytical capability and able to clearly articulate the critical role of data analysis in improving future performance. Without a convincing vision of how analytics will contribute to improving performance, there will be little buy-in from the coaches, and without the buy-in of the coaches, the project will never succeed. Analytics must be coach-led, with the coaches setting the agenda for the analyst whose responsibility is to provide rigorous and relevant evidence in a timely fashion to inform coaching decisions. Coaches must also be fully involved in designing the data collection system given that the most significant data are expert data involving the interpretation of performance relative to the tactical and technical expectations of the coaches.

The effective use of analytics in high-performance sport also requires greater integration of both data storage and data usage. In the early stages of an organisation's development of an analytical capability, however, it is common for data collection and analysis to be fragmented across individuals and functions. This leads to the creation of data silos with specific individuals collating and analysing data in their own spreadsheets quite independently of others with limited interactivity. Yet some of the most important questions involve the relationship between different types of data, for example, the relationship between count data on player contributions and tracking data on physical performance. This requires an integrated database that houses all the different types of data being collected and analysed within the sporting operation. But effective integration requires not only the physical integration of data, but also greater cross-functional coordination of the analysts, with closer working relationships between performance analysts—both qualitative video analysts and quantitative data analysts—and sports scientists.

Conclusions and Outlook

We have reached the stage in many high performance sports where the capability to collect data far exceeds the capability to use data. This "capability gap" is partly due to the explosive acceleration of data collection capabilities as a consequence of the digital revolution; however, it is also clear that cultural barriers have contributed by restricting the adoption of analytics by coaches. It follows, therefore, that coach education plays a key role in closing the capability gap. For example, modules on sports analytics need to become a core component in coach education programmes in all sports. Just as importantly, there also needs to be better sport-specific training opportunities for data scientists who want to work in high performance sport. Universities and colleges also need to offer sport-specific courses for data scientists akin to those now offered to performance analysts and sports scientists.

Looking forward, three analytical challenges face high performance sport. First, ass already discussed, the greater availability of tracking data on player positioning is opening up a mass of questions on the spatial aspects of invasion-territorial sports which will create a demand for analysts with a strong background in geometry and trigonometry. Second, as more "24/7" data is collected on individual athletes' competitive performances, training and general lifestyle, then there will be greater demand to use this data to design personalised training and lifestyle programmes to optimise competitive performances. These challenges will put a real premium on analysts with a multidisciplinary background, able to work with tactical, technical, physical and psychological data. The final challenge, related to the increasing diversity of the data available on individual athletes, is the increasing need for integrated real-time control systems to analyse and synthesise the totality of the data being collected so that significant changes in the likelihood of under-performance or injury can be flagged up and acted on immediately.

In summary, technology has transformed the data collection process in high performance sport. As a consequence, it is increasingly likely that the most successful athletes and teams will be those who are supported by a strong analytical capability that exploits data analysis to help create a winning margin.

Further Reading

Hughes, M., & Franks, I. M. (Eds.). (2008). The essentials of performance analysis. London, United Kingdom: Routledge.

A very comprehensive introduction to the many aspects of performance analysis. The chapter by Liebermann and Franks is a very useful survey of technological developments in data collection in high performance sport.

Schwartz, A. (2004). *The numbers game: Baseball's lifelong fascination with statistics*. New York, NY: St. Martin's Press.

This book tracks the history of statistical analysis in baseball and, as the subtitle suggests, shows that the interest in analysing baseball data is almost as old as the sport itself. Read this before reading Moneyball to properly appreciate the antecedents of sabermetrics use by the Oakland A's.

Lewis, M. (2003). Moneyball: The art of winning and unfair game. New York, NY: Norton.

This book was a real game-changer in high performance sport, providing a success story to inspire other teams to emulate the A's in the use of analytics. The author, Michael Lewis, was a financial trader before becoming a writer and has a real empathy with how Billy Beane used analytics to exploit misperceptions of market value. Read the paperback edition published in 2004 for an afterword on the hostile reaction of the baseball industry to the book.

Carroll, B., Carroll, B., Palmer, P., & Thorn, T. (1988). The hidden game of football. New York, NY: Warner Books.

One of the earliest attempts to provide a detailed statistical analysis of an invasion-territorial sport, in this case American football. They argued for the importance of the situational context—score, field position and time remaining. At the core of their analysis is the valuation of different plays based on the impact on win probabilities. The authors were one of the first to question the conventional wisdom of kicking on fourth down.

Anderson, C., & Sally, D. (2013). The numbers game: Why everything you know about football is wrong. London, United Kingdom: Viking.

A very readable account of the state of play of analytics in association football. It contains a considerable amount of data analysis interspersed with insights from analysts working in the professional game. The authors cover a variety of topics including the increasing rarity of goals, the undervaluation of defensive play, the importance of possession, the optimal timing of substitutions and the performance effects of sacking the manager.

References

Cook, E. (1964). Percentage baseball. Cambridge, MA: MIT Press.

Davenport, T. H., & Harris, J. G. (2007). Competing on analytics: The new science of winning. Boston, MA: Harvard Business School Publishing.

Evers, J. J., & Fullerton, H. S. (1910). Touching second: The science of baseball. Chicago, IL: Reilly & Britton.

Gerrard, B. (2007). Is the Moneyball approach transferable to complex invasion team sports? International Journal of Sport Finance, 2, 214–230. Retrieved from http://www.fitinfotech.com/IJSF/IJSF.tpl

Hughes, M., & Franks, I. M. (Eds.). (2008). The essentials of performance analysis. London, United Kingdom: Routledge.

Lewis, M. (2003). Moneyball: The art of winning and unfair game. New York, NY: Norton.

Mills, E. G., & Mills, H. D. (1970). Player win averages: A computer guide to winning baseball players. New York, NY: Barnes.

Rickey, B. (1954, August 2). Goodbye to some old baseball ideas. Life, 79–89.

Scully, G. W. (1974). Pay and performance in Major League Baseball. American Economic Review, 64, 915–930. Retrieved from https://www.aeaweb.org/aer/index.php

Schwartz, A. (2004). The numbers g*ame: Baseball's* lifelong fascination with statistics. New York, NY: St. Martin's Press.

Shea, S. (2014). Basketball analytics: Spatial tracking. Marston Gate, United Kingdom: Author.