Journal of Applied Sport Management

Volume 14
Issue 4 *The Sports Analytics Revolution*

Article 3

12-1-2022

Sport Isn't Sacred and Analytics Isn't New: Challenging Common Notions About Sports Analytics

Grant Morgan
Baylor University, grant_morgan@baylor.edu

Marshall Magnusen
Baylor University, Marshall_Magnusen@baylor.edu

Recommended Citation

Morgan, Grant and Magnusen, Marshall (2022) "Sport Isn't Sacred and Analytics Isn't New: Challenging Common Notions About Sports Analytics," *Journal of Applied Sport Management*: Vol. 14 : Iss. 4.
https://doi.org/10.7290/jasm14q8ff
Available at: https://trace.tennessee.edu/jasm/vol14/iss4/3

Journal of Applied Sport Management
https://doi.org/10.7290/jasm14q8ff
Vol. 14, No. 4
https://trace.tennessee.edu/jasm/vol14/iss4/3

Sport Isn't Sacred and Analytics Isn't New:

Challenging Common Notions About Sports Analytics

Grant Morgan
Baylor University
Marshall J. Magnusen
Baylor University
Please send correspondence to: Grant Morgan, Grant_Morgan@baylor.edu

---

Sports analytics, which can be summarized as the application of mathematical and statistical principles to sports to enhance athletic performance and create a competitive advantage, went mainstream in the sport management discipline after the popularization of the book, *Moneyball*, by Michael Lewis. His book, published in 2004, detailed how Billy Beane, then general manager of Major League Baseball's (MLB) Oakland Athletics, used various statistical approaches to identify talent and create a competitive advantage for his team while using a small budget. Considerable debate has taken place about the benefits and risks of such information since the book was published, and Big Data (i.e., data sets that are too large and/or too complex to be easily managed with traditional data-processing software) and analytics have become embedded across collegiate and professional sports.

There is no shortage of opinions on the role analytics should play in sports. From academics and sport journalists to coaches, athletes, and sport enthusiasts, opinions run the gamut from analytics are the future of sports to analytics are the end of sports. For example, Ben Shields, a senior lecturer at MIT Sloan, has argued that analytics can be helpful to teams and that data-driven decision making represents a union between science and art. On the other end of the spectrum is Charles Barkley. Barkley, an 11-time National Basketball Association (NBA) All-Star, quipped that "analytics is crap" and that analytics supporters are "a bunch of guys who ain't never played the game [and] they never got the girls in high school" (Golliver, 2015). Of the same mind as Charles Barkley, writer Phil Mushnick of the New York Post argued that analytics are ruining baseball because it has transformed The Game into a home run-or-strikeout processional. He lamented that "Everyone sees it, everyone knows it, everyone recognizes it is killing baseball, yet it persists. Baseball is ruled by math based on the broadest possibilities, ones that spread corrosion posed as wisdom" (Mushnick, 2022).

Ultimately, the matter of analytics and whether they hurt or help sports is not straightforward; it is nuanced. However, the need for nuance has often given way to sport fandom and nostalgia on one side and a passion for data and analysis trumping all forms of coaching intuition on the other. Thus, though the present continuum of opinions on sport analytics is interesting, often informative, and at times amusing, it is not complete. Refinement is needed. To provide such distinction to the debate about whether sport and numbers can cohabitate in modern day athletics, three areas are explored (albeit briefly) in the present paper. The first area focuses on the newness (or lack thereof) of analytics. The second area focuses the objectivity of analytics. The third area focuses on the idea that athletic competition is somehow sacred and should not be soiled by applying various statistical methods to practical sport performance problems. Each of these areas is discussed in the ensuing sections.

Analytics Isn't New

Professional sports, at the management-level, is about maximizing wins or maximizing profits. If you adhere to that simple dichotomy, then using analytics strategically to achieve one of these ends does not ruin sports, it enhances sports. In fact, sports teams have always used some sort of analytics to create a competitive advantage. Ostensibly, we would even engage the same processes "manually" in sports analytics as we do with machine-based tools, only much, much slower. Thus, the core problem does not rest with analytics, and data analysis is not suddenly a novel phenomenon in sports. The problem is that chosen methodologies and output may exceed our ability to explain the game, which as a consequence, can lead to feelings to that spirit of the game has been removed.

At the risk of sounding pedantic, a more nuanced definition of analytics would be helpful before moving forward. Over the past decade, as noted at the outset of this paper, the term "analytics" is commonly used and understood to refer to any number of advanced techniques and/or algorithms, such as random forests and machine learning, applied to data collected during and about sports. Underneath the complexity of these tools is a primary goal that has long been incorporated in the sports-related decision-making: prediction and classification via pattern detection. In fact, we can forgo a technical discussion of the origins and procedural details of the advanced analytic tools used in sports because they, in some form or fashion, share these goals.

Scouts, coaches, trainers, managers, athletes, and fans have all engaged these underlying functions for as long as they have sought strategic and competitive advantages…certainly long before the uniquity of powerful personal computers and proliferation of cloud computing and parallel processing. In short, myriad sports stakeholders have attempted to detect patterns in situations and circumstances that preceded various outcomes. The "novelty" of analytics is simply the application of tools long used in computer science, statistics, and other quantitatively oriented disciplines. Many have tried to determine which combinations and/or sequences of events and/or characteristics came before some important or consequential outcome. The process is not unlike a pitcher having a very slightly different delivery motion for a fastball versus a slider; the delivery motion can be used to predict pitch type. The process is also not unlike the star classification system for high school and collegiate athletes to express the likelihood of success at a higher level of competition; certain skills, experiences, physical attributes, etc. are associated with competitive success.

Unfortunately, simple and advanced analytics alike appear to be falsely viewed as deterministic rather than probabilistic such that even a single counterexample can, at times, result in stakeholders wishing to abandon "the analytics." One benefit of statistics-based tools is their provision of conditional outcomes and probabilities. Such results can inform not only which decisions to make, but also how much confidence to place in such a decision, considering the evidence. In the next section, we address a second fallacy of the use of analytics in sports; that is, analytics is not objective at its core.

Analytics Isn't Objective

Analytics is perceived by many as removing the human aspect from sports, but this is not at all true. This perception of analytics may be caused by many people misunderstanding the distinction between objectivity and subjectivity. The relationship between subject and object has been fodder for philosophers for centuries, and we do not wish to engage a dialectic on the matter here. For our purposes, we simply contend that observation is selective and an inherently subjective process. Popper (1963) provided a classic account from his physics class 25 years earlier in which he instructed his students to "Take pencil and paper, carefully observe, and write down what you have observed." To his apparent delight, the students asked what they were to observe exactly, which makes his point clear. Observations occurs when we direct our senses and design tools to collect certain data *selectively*. As such, the information used in analytics has been assigned some value, implicitly or explicitly, by humans. Consider how wind speeds, amount of drop or break, revolutions and so on are recorded about every MLB pitch. How come? Why is the street number of the umpire's mailing address or the number of unique words in the catcher's favorite song not recorded as well? For these examples, the atmospheric and pitch-related variables are collected, and the other variables are not collected. The reasoning for data collection is likely based on some sort of guiding thought and expectation for which variables do (not) or may (not) have some utility in understanding the game.

Clearly, the variables that are collected, recorded, and analyzed are chosen by people. We use computers and analytics as tools to execute the programs and analytic routines we have provided. For all intents and purposes, computers do not have value systems that were assigned to them. Sticking to sports, rather than deviating to a discussion of whether artificial intelligence (AI) can truly create its own value system, our point is this: sports analytics does not remove the human decision-making process. However, though humans assign a value system, that does mean that the models that are generated can be well interpreted or explained.

Not all collected information have the same salience and thus, potential usefulness. When information becomes used for a purpose, it becomes more than data; it becomes evidence. Unlike data, "evidence can be epistemically modeled or explicated by a proposition in a deductive logical system" (Stroing, 2022, p. 422). Epistemological models, explanations, propositions, and deductive logic systems are all created by people. Treating analytics as an objective enterprise is like piling raw meat on a countertop to be grabbed in handfuls and stuffed into a meatgrinder in hopes that the result is not only a "dish," but a delightful dish that is superior to the dishes ground by competitors.

The difficulty humans have with probabilistic thinking, particularly conditional probability, is both well-documented and clearly evident in the numerous logical fallacies we have identified to explain the phenomena. Common logical fallacies associated with analytics are ecological and exception fallacies. The ecological fallacy occurs when researchers base conclusions about individuals on the analysis of grouped or aggregated data. The exception fallacy occurs when researchers draw conclusions about groups based on the characteristics of one member of the groups. In many instances, these fallacies manifest in the form of availability bias whereby humans more readily recall examples of certain outcomes than nonexamples. For instance, someone may very likely overestimate the success frequency of half court buzzer beater shots because they are shown on highlight reels or simply because missed half court shots are just not very memorable. Some readers will no doubt recall the union and intersection rules from probability theory, but accurately computing the probability of some outcome conditioned on tens, hundreds, or even thousands of other variables is only feasible using a computer.

Although we do rely on algorithms and machines to identify complex patterns and return highly conditioned probabilities, two truths remain: (1) analytics are only consistently useful to the degree that we can understand the output, and (2) the results are probabilistic. Analytics return results that may apply in the long run. Not every fourth down attempt will be converted. Not every selected designated hitter will bat in a run in a particular situation. Not every five-star recruit will be more successful than every three-star recruit. Flipping a coin may have yielded the same outcome in a given instance. To point out the obvious, there are countless facets and variables whose existence we do not and cannot account for, much less whose individual and combined effects can be incorporated in an explanatory way into an impossibly complex set of dynamic relationships. Computers can be used to iterate billions of possible outcomes and combinations of variables in a simulated environment to determine the conditional likelihood of some outcomes. But in reality, only one decision is made. Walk a batter with no outs and a runner on second or pitch to the batter? A choice is made and we cannot know what would have happened had we made a different decision (i.e., counterfactual). Until we can access some parallel timeline where the counterfactual exists, we must compare results and outcomes against a guiding explanatory framework in order to determine (1) the extent to which using data as evidence is justified and (2) the utility of an analytic model. Ultimately, neither process is objective because machines should be used to make analysis easier, not allow us to suspend our judgement (Kline, 2016).

Athletic Competition Isn't Sacred Ground

So far, we have contended that the underlying functions served by analytics are not new and that analytics does not really remove human input. Curiously, those who oppose the use of analytics or (to be fair) the prevailing role analytics plays in sports-related decision making often use these arguments to support their opposition. Suppose it were true that analytics via machine learning or highly sophisticated algorithms somehow fundamentally changed sports-related decision-making and completely removed human input from game strategy. Why should this disqualify them from use in sports? The onus remains on the athletes to execute a strategy. In fact, many uses of technology, such as instant replay, are intended to isolate the effect of the athlete(s) on the outcome. Does it not support the integrity of sport more to rule out the effect human error (e.g., incorrect calls, poor subjective scouting, etc.) may have on outcomes rather than athletic performance?

Few people complain about the use of AI and machine learning for advances in medicines and drugs used for injury/illness prevention and treatment. Rossi and colleagues (2018), for example, demonstrated how GPS measurements could be used for predicting sport injuries via machine learning in professional soccer. Through the power of machine learning, athletes can better know when to train and when stop, which is a very desirable outcome for preserving both the athlete and the sport. However, when it comes to competition, there seems to be a belief by many fans, athletes, and sport professionals that competition and strategy must be manual to maintain integrity. Outside of sports, analytic advantages do not pose a problem, but even within sport, few people object to the use of sophisticated analytics if used in, say, sports medicine to prevent an injury. We contend that, for the reasons introduced, the argument against the use of analytics of sports is illogical if the integrity of sports is, in fact, rooted in the successful execution of strategy via athletic performance.

Closing Thought

Sports analytics isn't new or objective; athletic competition isn't sacred ground. These were the three areas discussed in this paper. While exploration of these areas won't end the sports analytics debate, it is the hope of the authors that bringing them to the forefront of the conversation rather than the background, may spark critical conversations about how academics, athletes, consumers, and sport professionals can all come to better understand and more effectively evaluate the phenomenon.

References

Golliver, B. (2015, February 11). TNT's Charles Barkley rants about basketball analytics, jabs rockets GM. *Sports Illustrated*. https://www.si.com/nba/2015/02/11/charles-barkley-analytics-video-daryl-morey-houston-rockets-gm

Kline, R. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford.

Mushnick, P. (2022, April 21). *New York Post*. MLB's marriage with analytics ruining baseball for the extreme worse. *New York Post*. https://nypost.com/2022/04/21/mlbs-marriage-with-analytics-changing-baseball-for-the-worse/

Popper, K. (1963). *Conjectures and refutations: The growth of scientific knowledge.* Routledge.

Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernández, J., & Medina, D. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *PloS one, 13*(7), e0201264

Ströing, P. (2018). Data, evidence, and explanatory power. *Philosophy of Science*, *85*(3), 422-441.