

Changing the Game: “Dr. Dave” Schrader on Sports Analytics

by Walter Tichy

Editor’s Introduction

Dave Schrader, known to his friends as “Dr. Dave,” worked for 24 years in advanced development and marketing at Teradata, a major data warehouse vendor. I’ve known him since his days at Purdue, where I was on the faculty and he was a Ph.D. student. Although he retired in 2014, he still gives talks on business analytics and since retiring has spent time exploring the field of sports analytics. In the past two years, he has given more than 40 sports analytics talks on college campuses—often meeting with students, faculty, as well as athletic departments—in both the USA and Germany.

*Walter Tichy
Associate Editor*

Changing the Game: “Dr. Dave” Schrader on Sports Analytics

by Walter Tichy

Walter Tichy: Dr. Dave, what is sports analytics and what is it used for?

Dr. Dave: Sports analytics is the art and science of gathering data about athletes and teams for analysis to create insights that improve sports decisions, like deciding which players to recruit, how much to pay them, who to play, how to train them, how to keep them healthy, and when they should be traded or retired. For teams, it involves business decisions like ticket pricing, as well as roster decisions, analysis of each competitor’s strengths and weaknesses, and many game-day decisions.

It is a hot area right now, because sports is big business. According to Price Waterhouse, the entire sports industry generates \$145B per year worldwide, with an additional \$100B in legal and \$300B in illegal gambling. How much money is spent to make better sport decisions using analytics? Another study I found says investments in analytics will go from \$125M per year in 2014 to \$4.7B by 2021.

The analytics split nicely between the front-office and back-office. Front-office analytics include topics like analyzing fan behavior, ranging from predictive models for season ticket renewals and regular ticket sales, to scoring tweets by fans regarding the team, athletes, coaches, and owners. This is very similar to traditional customer relationship management. Financial analysis is also a key area, especially for the pros where salary caps or scholarship limits are part of the equation. Back-office uses include analysis of both individual athletes as well as team play. For individual players, there is a focus on recruitment models and scouting analytics, analytics for strength and fitness as well as development, and predictive models for avoiding overtraining and injuries. Concussion research is a hot field. Team analytics include strategies and tactics, competitive assessments, and optimal roster choices under various on-field or on-court situations. For a great overview of both uses of analytics at pro teams, I’d recommend Tom Davenport’s [article](#) [1].

WT: I recall a movie called “Moneyball,” in which a coach for a baseball team is strapped for money. He decides to employ computer-generated analysis to recruit new players. Can you comment on this?

DD: Yes, the story about Billy Beane and the amazing Oakland A’s turnaround—highlighted in the 2003 book by Michael Lewis and the movie starring Brad Pitt in 2011—certainly popularized the idea of how a general manager can use data and analytics to turn a losing team into a winner. In this story, the A’s used analytics to draft players who were able to get on base, compared to traditional measures like stolen bases or runs batted in. That insight provided a competitive edge in drafting great players overlooked by other, richer teams. And it worked: They made the playoffs in 2002 and 2003.

WT: What are some of the examples you’ve worked on?

DD: As part of my volunteer “job” helping faculty and students who are part of the [Teradata University Network](#) (TUN), we’re putting together collaborative projects between academic departments and athletic departments. For example, on the business side, two of our projects involve analyzing season ticket holders at the University of North Carolina at Greensboro and Wright State University. We met with the associate athletic directors in charge of ticket sales who haven’t had the time to take a look at predicting which fans will renew season tickets or how different approaches to bundling tickets for multiple sports might drive attendance. They were happy to provide data to the business school faculty so students could study these problems and make recommendations.

On the team operations side, we have an interesting project underway in which a defensive coordinator at the University of Dubuque supplied us with a complete season’s game film and annotations for each American football play in Excel. Our goal is to create better predictive models for what play the offense might run next, given the offensive formation, the number of yards to go, the location of the ball on the field, the opposing team’s history of tendencies, and many other factors. Two teams of graduate students at Oklahoma State University are analyzing this data and providing the coach with some new insights, which he will incorporate into game planning.

WT: Can you explain a simple analytic technique?

DD: A predictive model the students built uses a decision tree with four factors to predict whether the offense is likely to run or pass on the next play (see Figure 1). Using a training set of 540 plays, they split at the first level based on the offensive personnel formation (21,12, 30 ... indicates a formation with lots of runners and tight ends; 10,11, 22 ... indicates more pass receivers). For each split, you can see the probability of a run or pass play. At the second level split for high probability passing formations, the model splits on whether it's first or second down, or third or fourth down. At the third level, they refine first or second down projections, which depend on the score differential. For third and fourth down, the split is on distance to the first down. What's noteworthy is this model was built in only one week, yet when the students ran it on the test data set, they achieved 75 percent accuracy—not bad!

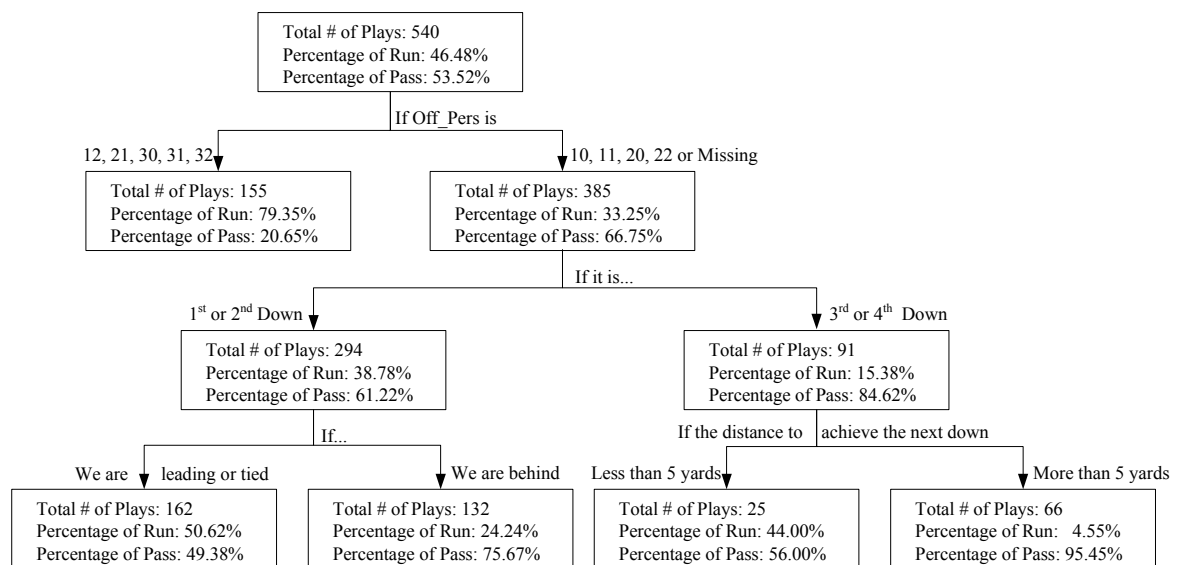


Figure 1. Decision tree for the likelihood of the offense to run or pass.
(Prepared by Nan (Peter) Liang, Oklahoma State University).

Another project at Bryant University in Rhode Island aims to help the football coach do better recruitment of players using three new predictive response time tests. You can view a [video](#) (“BSI: Sports Analytics - Precision Football”) of what we’re doing in this project.

WT: What data can be used to improve performance?

DD: There are many new market entrants for wearable devices, and all of these create data for performance improvements. All have strengths and weaknesses. For example, most GPS systems for player tracking work well outside but not inside. Some systems measure internal load (heart rate, respiration rate, blood pressure, etc.), while others use accelerometers to measure external load (athletic movements) and some do both. Much of this information helps trainers decide how much training is enough—athletes need a balance between stress and rest to achieve peak performance on game day. One of the best talks I’ve seen was by the head of sports performance for Cirque du Soleil, who explained how they work with choreographers and athletes to ensure the stress rates for new musical performances have peaks and valleys. They monitor load rates across multiple shows to decide when athletes need additional days of rest.

WT: And what about injury prevention?

DD: There are a lot of projects in this area, some sponsored by leagues or the NCAA (National Collegiate Athletic Association). There’s a great project underway at the University of Tennessee at Chattanooga and Auburn University to build predictive models for injuries based on some simple Android phone tests. We’re helping to recruit coaches and trainers to instrument their athletes on a variety of these tests so researchers can refine injury prediction models based on agility and reaction time data. This information might augment deeper load and training instrumentation from systems like Catapult and Zephyr, which many professional and some college teams use.

WT: Any other uses of data?

DD: Analytics is also coming on strong in sports journalism. At the University of Arkansas, a sports information reporter named Roland Liwag used a Zephyr vest to instrument their best basketball player. The coach’s goal is called “Fastest 40,” meaning he wants his athletes to be able to play at peak performance through a 40-minute basketball game without gassing. Just for fun, Roland put a vest on himself and then drew graphics that show how well Moses Kingsley was able to handle external loads (lots of running) while maintaining a constant internal load over 40 minutes. By contrast, the reporter put on the vest, and well, let’s just say he didn’t do so well.

Moses vs. Roland – No Contest

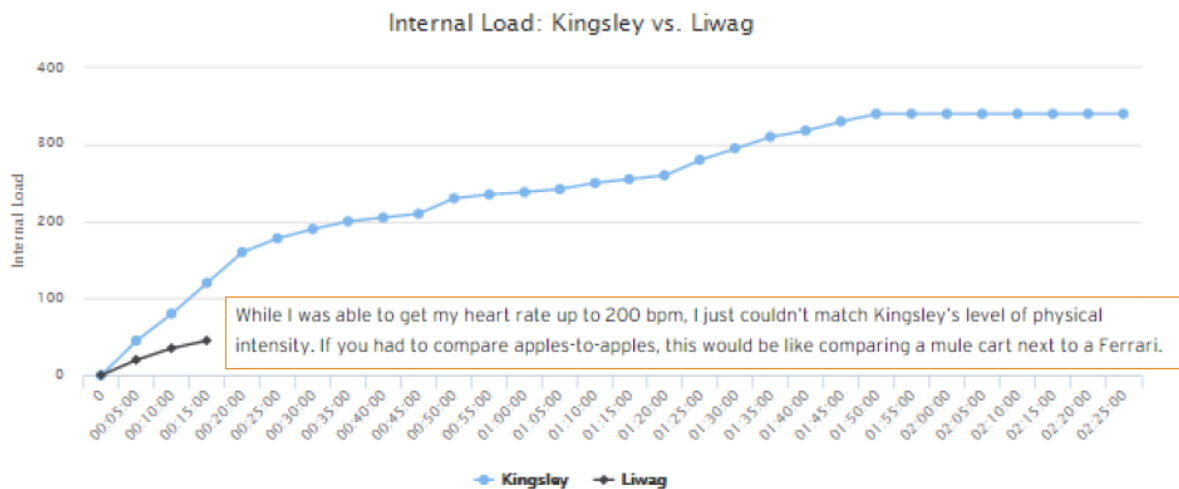


Figure 2. Fastest 40 Results.

(Source: University of Arkansas website; Data Hogs stories at <http://bit.ly/1T4X2bX>)

WT: Is there a difference in test outcomes between trained athletes and amateurs?

DD: One fascinating study I found in a TEDX talk by Prof. Jocelyn Faubert at the University of Montreal is called the “Motion Object Test” with eight animated balls on a screen. Four are highlighted, then they test subjects to see if they can pick out those four after they rapidly move around the screen. If someone gets a test right at one speed, then they speed things up for the next test until they fail. The research found significant differences between professional, college, and non-athletes in the fields of ice hockey, rugby, and soccer (see Figure 3).

These results can probably be extended to other sports where monitoring multiple events is a key cognitive skill. For example, linebackers in American football must predict what’s happening based on the offensive formation, receivers in motion, the context of the game, and which way the line is blocking on every play. You can see this test in the “BSI” episode I mentioned earlier. Another place to see these kinds of tests is on the [BrainHQ website](#) by Posit Science.

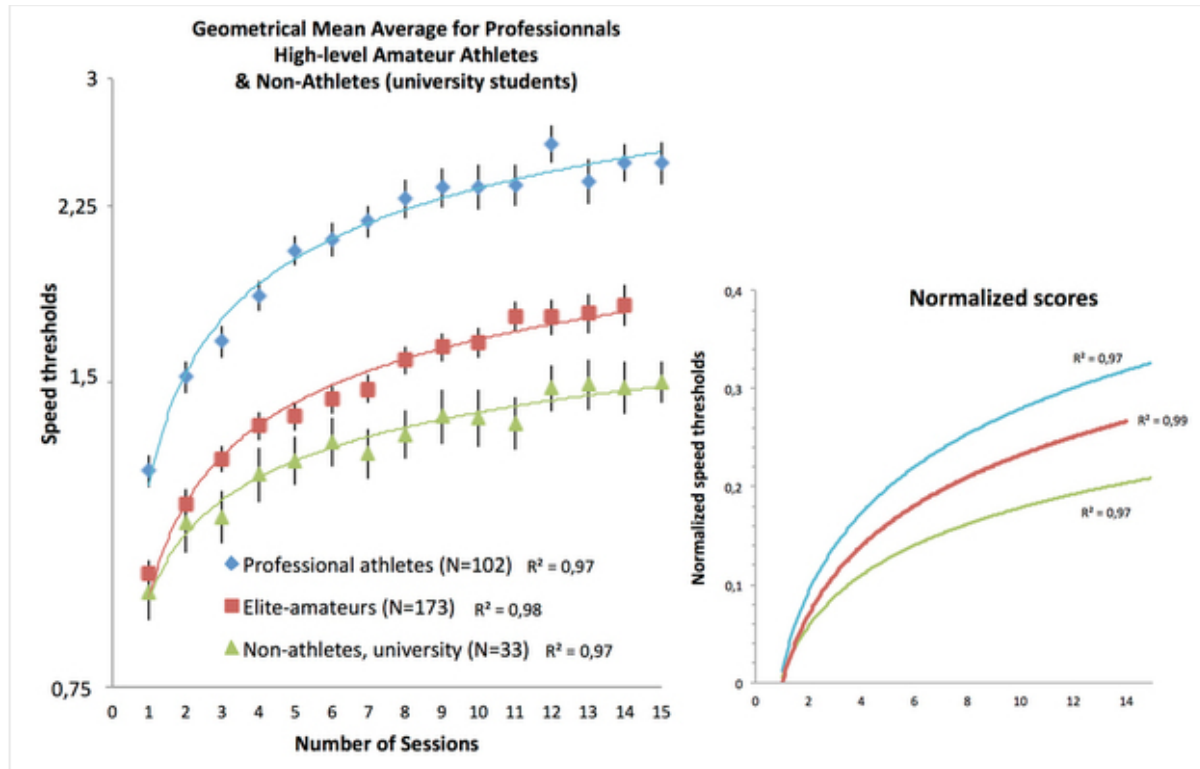


Figure 3. Geometrical 3D-MOT speed threshold means for 308 individuals on a log scale separated into professional, elite-amateur and non-athlete university students as a function of training sessions. (Source: Faubert, J. [Professional athletes have extraordinary skills for rapidly learning complex and neutral dynamic visual scenes](#). *Scientific Reports* 3, Article Number 1154 (2013).)

WT: And do sports analytics achieve better results?

DD: We’re beginning to see evidence that it works, but the proof will require the collection of data over time. AC Milan put together a lab back in 2002, measured up to 60,000 data points on each of their players, and reported a 90 percent reduction in injuries within the first year [1]. When you have star pro players who drive your revenue stream, keeping those people healthy and on the field makes a difference. Some teams also claim analytics are helping to extend the longevity of key players by not wearing them out. Achieving provable results can be difficult and is not unlike what I saw over years of interacting with businesses. The early-adopters want to get a competitive edge, so they do not openly say how big an advantage they are getting. The same holds true for the pro and college teams who are early-adopters.

WT: As long as only a few teams used analytics, it made a difference. Are we at the point where all major teams have to invest in analytics to get anywhere?

DD: Yes, interest is steadily growing. You can see that growing interest by observing which teams are sending attendees to the [MIT Sloan Sports Analytics Conference](#) (SSAC). I've gone to this amazing event in Boston the last two years. It has grown to 3800 attendees, and I counted 450 from professional teams. There was a big range of pro attendees team by team. Since it's in Boston, the New England Patriots had more attendees than anyone else, but 100 percent of the NBA basketball teams were in attendance, 80 percent of the NFL football teams, and 70 percent of the MLS soccer teams sent people to this event. And it's not just Americans; I met people from Australia, Germany, Italy, and Russia.

There's also been a big increase in the number of universities sending coaches, trainers, and students to this event. I even spoke with several high school math teachers who are now incorporating sports analytics into their curriculum. It's a good way to get students to study STEM subjects, because it all begins with data collection. There were also 25 vendors showing the latest products to instrument players and teams. So the bottom line is that teams are investing and growing their analytics skills.

WT: You talked about American football. What about soccer? What are meaningful measures for that game?

DD: Research in soccer is coming along at a fast rate. ACM held a Grand Challenge in 2013 with numerous college teams competing to analyze a mock soccer game that was played in Germany.¹ I liked this write-up by students at Vanderbilt University who participated in the contest [2]. The kinds of analytics included how far players were running and how fast, degree of ball possession by individual players and teams, and heat maps showing where on the field the ball was located. The focus of the student paper is on building a distributed event-based processing system to handle the real-time data and analytics.

While how far people run might be helpful for analyzing fatigue and driving game substitution decisions, the more interesting question is how each team plays offense and defense. An excellent piece of research can be found on the MIT SSAC website, which was presented this past March [3]. These researchers analyzed game play styles, showed who is good or bad at offense and defense for the English Premier League, and correlated game play styles with coaches (and changes of coaches).

¹ A description of the ACM DEBS 2013 Grand Challenge, including a link to data and required analytics queries, can be found at <http://bit.ly/1WOhkG2>.

For soccer fans, an excellent book that explodes many commonly held soccer myths is *The Numbers Game* by Chris Anderson and David Sally [4].

WT: Do you foresee that analytics will be used during games, perhaps to change tactics or replace players while the game is going on?

DD: That's a possibility, but often there are league rules that prevent the use of analytics during the game. Some leagues are experimenting with allowing sensors for measuring workload. I think this will evolve, but an interesting phenomenon is that many pro leagues want some degree of parity or at least table stakes for data and analytical insights. If one team has far superior insights because of their investments or vendor relationships (think Bayern Munich), and that pays off in a winning record, it forces the other teams to adopt technology. Thus, the competition for analysts will be high. In many pro leagues, the video systems and raw data are provided by the league to all the teams, but it's up to them to create their own deeper analytical insights.

WT: Wrapping up, how can people find more information about sports analytics?

DD: My top recommendation is to attend the MIT Sloan Sports Analytics Conference and take a look at all the research reports and video links associated with their website. Book early; it always sells out.

Faculty and students should take a look at the [Teradata University Network](#) (TUN). It's a free resource that provides the entire technical stack needed to do business intelligence research, including free licenses for the Teradata data warehouse, the Teradata Aster big data system, SAS statistics, Tableau visualizations, and MicroStrategy business intelligence tools. TUN is used in 114 countries by 2,500 universities. Five thousand faculty and 12,000 students have accounts, where they can access not only tools but also case studies, homework assignments and projects, and sample syllabi.

I've added a sports analytics section on the TUN website to highlight where people can go to find public sports data sets. We also provide a list of sample capstone or research topics that need to be explored, and make available some teaching materials so faculty can add interesting sports examples to their classes. I'm also developing compilations of things to read or view on the web, sport by sport. The website is run mostly by faculty for faculty—our Board of Advisors is international and the faculty come from the fields of management information systems,

marketing, and computer science. I'm especially interested in helping build out materials for computer science students, given my Purdue background. Feel free to contact me for details at drdaveschrader@gmail.com.

WT: Thanks for the tips, Dr. Dave!

References

- [1] Davenport, T. [Analytics in Sports: The New Science of Winning](#). International Institute for Analytics White paper. Feb 2014.
- [2] An, K., Tambe, S., Sorbini, A., Mukherjee, S., Povedano-Molina, J., Walker, M., Vermani, N., Gokhale, A., and Pazandak, P. [Real-time Sensor Data Analysis Processing of a Soccer Game Using OMG DDS Publish/Subscribe Middleware](#). Institute for Software Integrated Systems. Vanderbilt University. Technical Report ISIS-13-102. 2013.
- [3] Bojinov, I. and Bornn, L. [The Pressing Game: Optimal Defensive Disruption in Soccer](#). Presented at MIT SSAC 2016.
- [4] Anderson, C. and Sally, D. *The Numbers Game: Why Everything You Know About Soccer is Wrong*. Penguin Books, New York, 2013.

About the Author

Walter Tichy has been professor of Computer Science at Karlsruhe Institute of Technology (formerly University Karlsruhe), Germany, since 1986. His major interests are software engineering and parallel computing. You can read more about him at <http://ps.ipd.kit.edu/>.

DOI: 10.1145/2933230