Data Analysis Interface for USA Baseball Talent Evaluation Processes

In-Depth Analysis of Competing Beliefs Regarding Data Implementation in the Player Development Process

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Matthew Pezolt

October 27, 2022

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Professor Wayland, Department of Engineering and Society

Rosanne Vrugtman, Computer Science

Introduction

Baseball is in the midst of an identity crisis, 30 years in the making, over the role that data analysis plays in the player development process. The conflict turned mainstream with the publishing of *Moneyball* and the conflict of data analysis in sports really began from here (Lewis, 2003) (Burris, 2022). The professional ranks reluctantly have adapted to the availability of new data and research over time. However, since then, the battle has simply shifted from the professional to the amateur realm. Now that colleges have begun to embrace data analysis within their programs, as seen at Oregon State, the topic has become "how young can we start optimizing baseball performance?" (CDW). The desire to use advanced data analysis in the amateur/youth baseball industry has created a highly contentious debate between people who believe in new-age training methods and those who still vouch for traditional methods of baseball development.

While the implementation of advanced data analysis in amateur baseball is still a very taboo topic in baseball circles, USA Baseball was looking for a method to compare player data collected at their National Team Identification Events. USA Baseball works with hundreds of athletes all over the country each season. As a result, it is nearly impossible for all their National Team Identification Staff to attend each identification event, and even more difficult to see all the athletes being evaluated on a year-by-year basis, creating a need for a tool that could standardize the data collected from these identification events for more consistent analysis.

The debate surrounding the need for a tool like this and how it should be used within amateur baseball was a frequent topic of conversation between those of us there collecting the data and coaches who were curious about how the data would be used. Throughout the Summer,

I encountered beliefs spanning the spectrum, from coaches that love to use the data to coaches that think the data creates a crutch for youth players to lean on, providing a false sense of performance. For my STS Topic, I am hoping to evaluate the differing beliefs on the role data should play in the development of youth/amateur baseball players. Compiling beliefs from various sources, ranging from coaches, players, and player development personnel, I hope to come to a greater understanding of why this fundamental conflict is still so hotly contested in amateur baseball today, and how the conflict has contributed to the current state of the amateur baseball industry as a whole.

Technical Project: Data Analysis Interface for USA Baseball Talent Evaluation Processes

USA Baseball, the national body for United States representation in international baseball competition, wanted to further improve its player identification capabilities through enhanced data collection and visualization methods. USA Baseball fields five separate National Teams every Summer and those teams are as follows: The Collegiate National Team, composed of the best non-draft-eligible college players in the country. The 18U, 15U, and 12U National Teams, comprised of the best rising High School Seniors, 15-year-olds, and 12-year-olds respectively. Lastly, the Women's National Team, which is comprised of the best Women's Baseball players in the country, ranging from 16 to 41 years of age. The desire to send these teams into international competition and dominate drove the desire to have a more standardized data analysis tool in the roster selection process.

I designed a web-based database/interface that re-imagined the way USA Baseball's Talent Identifiers were able to analyze athletes' quantitative performances. The interface was developed using python and its library of data analysis packages, as well as its web-based

interface package called “Dash”. These packages allowed me to recreate mainstream data analysis tools already common in the baseball industry as well as the freedom to further expand and generate new visualizations where I felt popular analytical tools were failing to paint the entire picture.

Due to the number of players being evaluated that were already committed to high-level Division I College Baseball Programs, I determined that collegiate data analysis processes were the most applicable to what USA Baseball was trying to accomplish. Over the course of a week, I was able to dive into how college baseball teams handle data analysis, including the practices of the program here at the University of Virginia. After identifying the common tools and visualizations used in college baseball. I was able to develop a system that incorporated the industry standard visualizations such as hitter spray charts and pitcher pitch locations, while also developing new visualizations for areas of analysis that I felt the current state of data analytics in college baseball fell short of painting the entire picture, eventually encapsulating these visualizations in a web-based database containing the data collected from games that took place at the National Training Complex throughout the summer.

The project served as a great means to compare players in the national team selection process and served as an advanced preparation tool for evaluating competition in international tournaments. The project is still undergoing further enhancements as I think of new ideas for expansion, such as the calculation of more traditional baseball statistics. I will be deploying and testing this database with the University’s baseball program this season to further evaluate the long-term impact the project can have on player development at the collegiate level.

STS Research Topic: In-depth Analysis of Competing Beliefs Regarding Data Implementation in the Player Development Process

Player development, specifically in baseball, hinges upon the physical and mental maturation of athletes. As the players grow, they become physically more capable of performing at a high level, as well as more mentally capable of handling the trials and tribulations that come with participating in sports. Over the last decade, data collection and data analysis tools have become more accessible and have introduced a whole new variable into the world of amateur baseball training and development. The effects of this have separated trainers and coaches into one of two camps, pro-data, or anti-data. I hope to come to a better understanding of why these trainers/coaches feel the way they do, and what role they believe data analysis is playing in the development (both positive and negative) of amateur baseball players, both physically and mentally. Understanding the various viewpoints of various player development personnel will allow us to come to a deeper understanding of how the analytical boom in sports, specifically baseball, has changed both the amateur sports industry as a whole and the players partaking in it.

When diagnosing the mental maturation process of athletes, baseball provides one of the most interesting dynamics in all of sports. The game is built on failure. The best baseball hitters in the world are failing at a 70-75% rate. The psychological effects of such an environment have been studied at length, with plenty of case studies stemming from the realm of professional baseball. Those who believe in data-driven performance analysis of young athletes believe in the confidence created by being able to track quantifiable growth and the positive impacts this confidence can have on performance. Those who don't believe in these analytical methods believe that these measurables provide a false reality surrounding the athlete's performance, creating a crutch that holds the athlete back from being able to truly understand and alter their performance.

The physical development of athletes follows a trend that we have been seeing for more than a century. As the science behind training and athletics has evolved, the boundaries of what a human body is capable of have continued to grow. Athletes today are bigger, faster, stronger than we have ever seen them before, allowing for extraordinary physical feats at younger and younger ages. Looking at baseball, the effects of this can be seen in the college recruiting process. Players are reaching physical performance milestones, such as throwing a baseball 90 miles per hour, earlier and more often than ever before, leading to a recruiting system that has the top collegiate programs offering scholarships to athletes that are yet to enter high school. While some believe that these new performance heights are ultimately creating bigger and better opportunities for the athletes, others believe that new training regiments are too specialized and putting the athletes at greater risk of sustaining major injuries, and where a trainer/coach falls in their beliefs on this topic is often in-line with their stance on data-driven performance training.

Pitchers are generally believed in the industry to have the most to gain from specialized training practices, and these trainers who believe in data analytics and specialized training methods are following common research trends within the industry. Research indicates that the best way to improve pitch metrics is to optimize the kinematic efficiency of the pitching motion (Daneau et al., 2020). Research also shows that pitch spin rate is one of the leading indicators of pitching performance and that average pitch velocities have skyrocketed over the last decade (Higuchi et al, 2013), (Clair, 2018). Those who are proponents of new-age training believe that optimizing these metrics early in the physical maturation process will have the most benefit to players over the span of their playing careers.

Those who are generally opposed to this line of thinking cite growing injury rates in both professional and amateur baseball. Tommy John surgery rates have increased over the last

decade at both the professional and amateur levels (Arons et al., 2016). The root of these injuries stems from both the biological unnaturalistic nature of an efficient pitching motion, as well as overuse stemming from training specialization (Whiteley, 2007) (University of North Carolina). This collection of baseball trainers and coaches focus more on general athletic training at a younger age and waits to introduce more nuanced or baseball-specific features until an athlete has gone through a certain level of physical maturation. They maintain the belief that as an athlete grows, they become physically more capable of performing at a certain level and more mentally capable of understanding their own limitations.

Ultimately, over the span of my STS project, I hope to use current articles, as well as conduct interviews with current coaches, trainers, and players to understand the differing perspectives surrounding data analysis and data-driven training in amateur baseball. While I have laid out the problem as a two-dimensional black-and-white spectrum, I understand that some may be more polarized or more moderate than others. I hope to come to a deeper understanding of why those involved in baseball have engrained themselves in their beliefs, and how the new-found availability of data-tracking technology has impacted these beliefs. Have their beliefs been strengthened because of the new technology? Are they beginning to understand the other side? Then, using this, I hope to come to a better understanding of how these beliefs and mindsets have shaped the industry of amateur baseball as it exists today.

Conclusion

From my STS Research, I hope I can determine when in the amateur baseball circuit it becomes safe and responsible to use advanced data analysis on player performance. Combining that information with the database developed in the technical portion of my prospectus, I hope to develop a data analysis strategy for USA Baseball and UVA Baseball that can be implemented in

their player development and talent identification processes. If used correctly, I believe that this research can lead to meaningful impacts on player performance on the field, benefitting both the programs and the individual players themselves. For the programs, an increase in player development can lead to a better overall team, which in terms of USA Baseball can lead to more success in international competition, and for UVA Baseball, more success throughout a potential 70-game season. If the players can take advantage of this research, they could potentially see the dividends of higher in-game performance, leading to better opportunities and potentially a higher signing bonus when the MLB Amateur Player Draft rolls around. This research can be used as a guideline in the baseball industry, helping to establish when it becomes both feasible and physically safe to implement advanced data-driven training for amateur baseball players, leading to enhanced player performance and a better overall product on fields for all ages and skill levels.

References

Burris, T. (2022, March 14). *How data analytics have changed baseball forever*. Retrieved October 25, 2022, from https://onlinemba.montclair.edu/data-analytics-have-changed-baseball-forever/

CDW. (n.d.). *Data Analytics takes the field at Oregon State University*. CDW. Retrieved October 25, 2022, from https://www.cdw.com/content/cdw/en/articles/datacenter/data-analytics-takes-the-field-at-oregon-state-university.html

Clair, M. (2018, August 29). *Are pitchers really throwing harder than ever?* Cut 4. https://www.mlb.com/cut4/are-pitchers-really-throwing-harder-than-ever-c292153594#:~:text=not%20seeing%20things.-,Pitchers%20are%20throwing%20harder%20than%20ever.,an%20enormous%20boost%20over%20time.

Dunn, T. (2019, May 31). *Analytics and youth baseball: Data making its way into kids' game*. Retrieved October 25, 2022, from https://cronkitenews.azpbs.org/2019/05/31/youth-baseball-analytics/

Higuchi, T., Nagami, T., & Kanosue, K. (2013, January 23). *How baseball spin influences the performance of a pitcher - researchgate*. https://www.researchgate.net/publication/272467704_How_baseball_spin_influences_the_performance_of_a_pitcher

Hodgins J.L., Vitale M., Arons R.R., & Ahmad C.S. (2016, January 21). *Epidemiology of Medial Ulnar Collateral Ligament Reconstruction: A 10-Year Study in New York State.* The American Journal of Sports Medicine. https://journals.sagepub.com/doi/10.1177/0363546515622407

Lewis, M. M. (2003). *Moneyball: The art of winning an unfair game*. W.W. Norton.

Mercier, M.-A., Tremblay, M., Daneau, C., & Descarreaux, M. (2020, February 12). *Individual factors associated with baseball pitching performance: Scoping review*. BMJ open sport & exercise medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7047480/

Pennington, B. (2012, March 12). *Young Arms and Curveballs: A scientific twist*. The New York Times. Retrieved October 25, 2022, from https://www.nytimes.com/2012/03/12/sports/baseball/debate-grows-over-how-to-protect-young-pitching-arms.html#:~:text=For%20decades%2C%20it%20has%20been,joint%20than%20arms%20can%20handle.

Reinold, M. (2015, January 29). *Early Sport Specialization: The good, the bad, and how to fix it*. Champion Physical Therapy and Performance. Retrieved October 25, 2022, from https://championptandperformance.com/early-sport-specialization-good-bad-fix/

University of North Carolina. (n.d.). *Learning curve: UNC Arm Safety Study .PDF*. Docslib. Retrieved October 25, 2022, from https://docslib.org/doc/1822292/learning-curve-unc-arm-safety-study-pdf

Whiteley, R. (2007, March 1). *Baseball throwing mechanics as they relate to pathology and performance - A Review*. Journal of sports science & medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3778685/