Investigating MLB Draft Outcomes, 2002-2005

By Douglas Wachter

For consideration for the 2012 Jack Kavanagh Memorial Youth Baseball Research Award

[redacted]

[redacted]

[redacted]

Senior, University of Michigan

Literature Review

Each year, MLB teams select upwards of 1500 ballplayers from the amateur ranks in the Rule IV Draft, hoping that a small fraction of those players will go on to stardom at the Major League level. With bonus totals climbing each year, and top picks routinely receiving multimillion dollars even as teenagers, savvy teams seek to maximize the chance of selecting the players most likely to yield success over their long-term development. The purpose of this study is analyze the state of the draft from 2002-2005, and attempt to make inferences based on the trends in this data which could be of value to a team in developing an optimal draft strategy. Each draft year included in this study comes from the past decade, increasing its relevance to current and future drafts. While the changes to the draft as a result of the new Collective Bargaining Agreement are massive, these data can still assist teams in examining trends to help predict which players will develop most effectively and produce the most value over time.

Previous draft research from several prominent baseball statisticians informs my research into the Major League Baseball (MLB) draft. Foremost among these is the work of the father of draft research, and creator of Sabermetrics, Bill James. Mr. James is currently a Senior Advisor on Baseball Operations to the Boston Red Sox, and his Baseball Abstract annuals were the first public source of advanced statistical data in baseball.

In 1984, Mr. James turned his eye to the MLB draft. Looking at the history of the draft from 1965 to 1983, James reported that college players were much more valuable than high school players. In James' data, college players produced 84% more position-adjusted value than high school draftees. James also found that pitchers provided less value than position players, especially in the first round of the draft, and that players from California and the Northern states

produced a higher than expected return as compared with draftees from the traditionally more-heralded South.

The present study analyzes the relative value of college vs. high school players and value by positions, exploring whether James' findings hold true with data from more recent drafts. The scope of this investigation does not include the draftees' home states or any other geographical data, but if this study were to be expanded, analyzing these data would be a logical next step.

Studies utilizing data from the mid-1990s to the present suggest a change in the relative value of high school versus college players since the publication of James' seminal study. In a 2006 study using data from 1984-1999, Rany Jazayerli found that teams valued high school players more highly as draft choices beginning in 1995. In his conclusion, he extends his draft data through the 2005 season, and finds that from 1995-2002, the majority of players signed in the first 100 picks were selected out of high school. Jazayerli found that after 2002, and coincident with the publication of Moneyball, the trend reversed; college players once again became much more popular picks than high school players. Jazayerli suggests that this trend is related to other MLB teams' adoption of the Oakland Athletics' data-fueled draft strategy, which dictated that team's strong preference for college players.

This study will make an effort to determine whether the market skewing toward college players represents a new inefficiency, in that equally skilled high school players are drafted later, making high school players more valuable picks. Though Jazayerli's 2006 study did not allow him to analyze the pro performance of players drafted in the early 2000s, we can now expand on his research by doing just that.

Additionally, Victor Wang's 2009 study, using data from 1990-1997, explored draftees in terms of their value as compensation picks for the loss of free agents. Wang used a somewhat

similar methodology to the within study, although Wang analyzed a player's first six years in the majors rather than his first three, and limited his study to the first 100 picks in the draft. Wang also looked at positional valuations and his findings are in conformance with Bill James' earlier study, in suggesting that hitters continued to be more valuable picks than pitchers throughout the 1990s. The conclusions of Wang's study provide another more contemporary analysis of how MLB teams today can maximize the value of their picks.

Research Aim

Overall, past research suggests that college players provide more value as picks, and that hitters are more valuable than pitchers. In order to more closely analyze these distinctions, this study: 1) separates college players from high school players at each position, to determine whether certain groups of position players are more valuable than others within age groupings, and 2) analyzes the differences in the expected production of high school and college players at any given position. This study also reviews and compares the findings of the James, Jazayerli, and Wang studies and places them in the context of data from the most recent decade, and uses updated methodology, to determine how closely the trends hold. Jazayerli ends his extensive, 12-part study with his "overarching Golden Draft Rule": "The only rule that isn't subject to change over time is that all the other rules are."

In the spirit of Jazayerli's injunction, I hope to characterize these "rules," analyze trends since the beginning of the 21st century, and draw inferences as to how teams might behave in order to maximize the long-term output of their draft class. Statistics are a powerful tool to understand many aspects of the world around us, but they need to be used carefully and with an understanding of the implications of the context from which the data is derived. For most tests, this factor is taken into account in setting the desired confidence level for the test, which is

essentially a measure of how sure we are the results of our statistical testing accurately reflect the condition of the population we are studying. For example, in medicine, the result of a test might be the release of a new drug to market, so a study's confidence level is set at 95% or often higher, as the release of a bad drug could be unhelpful at best, and dangerous at worst, for those who need the treatment. In baseball, unfortunately, we often don't have the luxury of that level of certainty. In Jonah Keri's "The Extra 2%," Keri quotes Rays owner Stuart Sternberg in the prologue: "We've worked hard to get that extra 2%, that 52-48 edge." (Keri, 13) Because the market for major league talent is so competitive, a team will rarely be able to negotiate a deal they can be completely confident will be successful. However, we can use statistics to isolate trends that help us gain some understanding of a player's likely ability to develop. In amateur scouting, that understanding will continue to largely be driven by the assessments of experienced talent evaluators, but statistics can help us frame the expertise of those scouts and better understand the context in which they are delivered.

Methodology

In this study, I chose to use all players who signed after being drafted in the first 10 rounds of the 2002-2005 MLB Rule IV Drafts. This yielded a total pool of 1,179 draftees, including data for each player's draft class and selection number, team, position, age, school, and highest level reached in organized baseball. I then inserted into the data for all players who reached the Majors, each player's production in Fangraphs' Wins Above Replacement (WAR) in his rookie season in the Majors, and for two subsequent years. This omnibus dataset provides the best proxy for the true Major League value of draftees, rather than simply relying on the percentage of each type of pick that reaches the majors or another less reliable estimate of value.

Two clarifications should be made about positional coding. Players are coded based on the position at which they were drafted. For example, both Ryan Braun and Alex Gordon have developed into star-level left fielders, but because both were drafted at third base they should serve to increase teams' confidence in drafting third basemen, potentially with the knowledge that a move to left field could be in the player's future if the hot corner doesn't work out defensively. Additionally, while I certainly appreciate the difference in value between the taxing, up-the-middle defense required of a center fielder and the less strenuous defensive responsibilities of a corner outfielder, I was not confident enough in the classification of different outfield positions to simply use the raw data and felt it would be more prudent to group all three positions under the umbrella of outfielders.

Assumptions

This study uses data from 2002-2005 because, after careful consideration, several factors led to the conclusion that this pool is neither too recent nor too old. If the data were too recent, the study would miss data from players who are likely to produce in the Majors in the future. For example, at the outset of the analysis the data included the 2006 draft, as well as the years that were ultimately used. Looking into the 2006 draft, however, I determined that many players likely will not have produced the majority of the value that they ultimately will for their drafting teams. The example that made this obvious to me was Toronto pitcher Kyle Drabek, a pitcher drafted out of high school who earned his first extended stay in the Majors last season. While he's had some developmental setbacks, Drabek is still a promising prospect, and is likely to produce value in the next few years at the Major League level. I believe that, like Drabek, many prospects from the 2006 draft have yet to give their teams the three years of cost-controlled production that those teams will eventually receive, meaning that production likely exists that

would not be considered in our study. This future production is also likely skewed heavily to high school players, who are obviously much younger when they are drafted and therefore often require much more time in the Minors before they are prepared for MLB action. Avoiding the 2006 draft helps to avoid this factor, which would result in the study overstating any advantage college players have over high school players as compared with the true difference in talent levels between the two groups.

By using data beginning in 2002, this study also analyzes data that is recent enough to be applicable to today's draft, at least to some extent. The market for draftees is constantly changing, and in order for these results to be relevant to teams' draft strategies, using the most recent possible data is imperative. Since all data in this study is pulled from within the last decade, this study is more pertinent to current and future drafts than previous literature.

The study uses the first ten rounds of the draft because those draftees are by far the most significant, both in terms of the resources devoted to them by their teams and in terms of the players' ability. Many players selected later in the draft decide not to sign, but rather, return to school. Later players' skill levels are usually lower than players drafted earlier, and earlier players get much larger signing bonuses as well. Exploiting inefficiencies in the early rounds of the draft will result in larger advantages for teams, as there's simply better talent available. I also believe using the first ten rounds produces a large enough sample to create data with some predictive value, as studies such as Wang's research may be limited by their smaller scopes. By using roughly three times more draftees per draft than Wang, this study is able to consider less years overall. As a result, this study is condensed to study only four drafts, allowing the use of data that is as recent as possible.

This study uses players' Fangraphs Wins Above Replacement (WAR) total in their first three years. WAR aggregates a player's production from each phase of the game, including batting, fielding, and pitching. This is the best total-value statistic currently publicly available to sabermetricians, and using it allows us to compare players with different skill sets; for example, position players who derive most of their value from their defense to those whose value comes from their ability with the bat.

Next, using the first three years of a player's career is significant for two reasons. First, in a player's first three years, he is completely cost-controlled to his team; straight out of the Minors players receive the Major League minimum salary, which in 2012 was $480,000 per year. After a player's first three years, that player becomes arbitration-eligible, meaning his salary is tied much more closely to his production. In a player's three arbitration seasons, studies suggest that that player's salary averages 40, 60, and 80% of his open market value. As the player becomes a larger financial imposition to his team, he becomes much more likely to be traded or non-tendered. In addition, that player's production no longer comes essentially for free, as it did in his first three seasons, meaning that player's total value to his team is reduced significantly. Using only three years, rather than six, also means that players need much less Major League experience to provide a full set of data for this study. This allows the use of much more recent draft data than earlier studies on the subject.

Results

1. College draftees are not uniformly superior to high school draftees.

2. Positions near the center of the defensive spectrum (OF, 3B) average the most pre-arbitration production overall.

3. College draftees average more pre-arbitration production than high school draftees at defensively demanding positions (2B, SS, 3B).

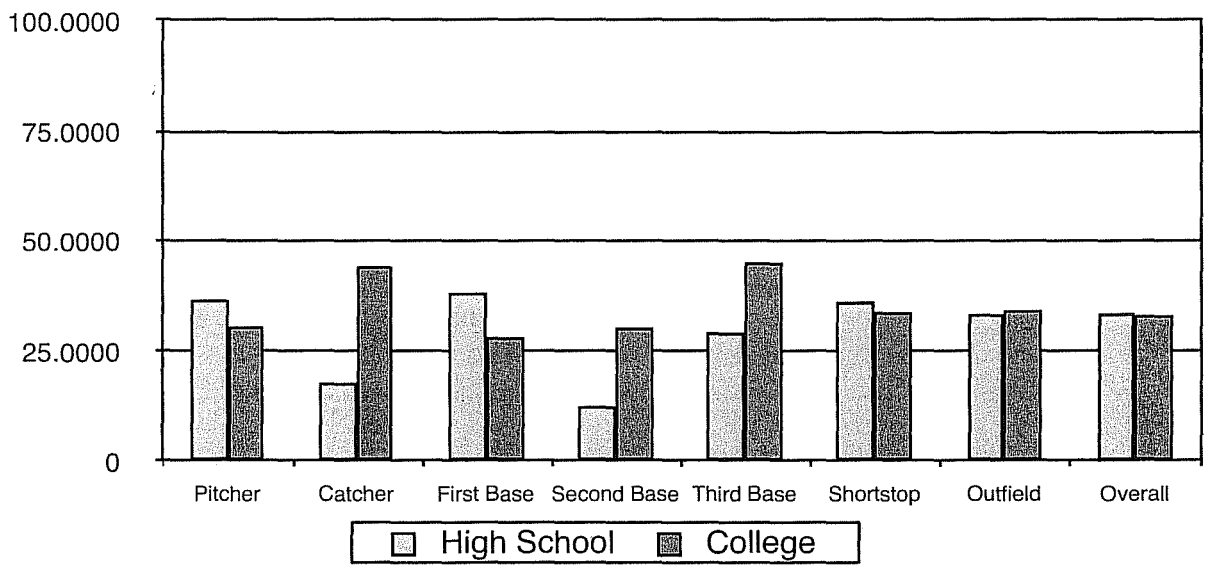
4. High school draftees average more pre-arbitration production than college draftees at less defensively demanding positions (1B, OF) and as pitchers.

Data Tables

Position	Draftees	High School Draftees	High School Major Leaguers	College Draftees	College Major Leaguers
Pitcher	602	178	65	424	130
Catcher	87	28	5	59	26
First Base	60	21	8	39	11
Second Base	41	8	1	33	10
Third Base	64	24	7	40	18
Shortstop	101	36	13	65	22
Outfield	224	78	26	146	50

Table 1: Raw total numbers of draftees by position; total numbers of draftees at each age level; number of draftees at each position/age level with sufficient Major League service to lose rookie eligibility. (130 ABs or 50 IP)

Figure 1: Percentage of Draftees Losing Rookie Eligibility by Age and Position



Position	High School Draftee MLB Rate	High School Draftee AVG MLB WAR	College Draftee MLB Rate	College Draftee AVG MLB WAR
Pitcher	36.5%	0.76	30.6%	0.52
Catcher	17.9%	0.50	44.1%	0.43
First Base	38.1%	1.0	28.2%	0.22
Second Base	12.5%	0.04	30.3%	0.89
Third Base	29.2%	0.90	45%	1.62
Shortstop	36.1%	0.25	33.8%	0.89
Outfield	33.3%	1.17	34.2%	1.01
Overall	33.5%	0.78	33.1%	0.68

Table 2: Percentage of each draftee category that reached the Major Leagues and lost rookie eligibility, and expected value for a draftee at each position in their first three years at the Major League level.

Position	Draftee Expected Value (WAR)
Third Base	1.35
Outfield	1.06
Second Base	0.72
Average	0.72
Shortstop	0.66
Pitcher	0.59
First Base	0.50
Catcher	0.46

Table 3: Positions ordered by expected value.

Position	High School Expected Value (WAR)	College Expected Value (WAR)	High School- College Expected Value (WAR)
First Base	1.0	0.22	0.78
Pitcher	0.76	0.52	0.24
Outfield	1.17	1.01	0.16
Catcher	0.50	0.43	0.07
Shortstop	0.25	0.89	-0.64
Third Base	0.90	1.62	-0.72
Second Base	0.04	0.89	-0.85
Overall	0.78	0.68	0.1

Table 4: Positions by largest gap between expected production for high school and college draftees.

Discussion

Comparison of Age Groupings

I first analyzed the data to test the sweeping, age-based pronouncements that are commonly touted in the media. As draft day nears, a flurry of articles proclaim that the value of college players exceeds high school players. Unsurprisingly, the data actually reveal that college players are not uniformly more valuable than high school players; assuming that they are is far too simplistic an observation on which to craft an efficient draft strategy. High school and college players, in fact, provide similar value overall, with an expected value of .784 WAR for high school draftees and .683 for players signed out of college for a statistically insignificant 0.1 WAR difference (Table 2). However, some positions seem to better lend themselves to an early, post-high school start to a professional career, while others are likely to produce more value after another level of amateur development in college.

While teams do consider those trends in their draft picks, overall their selections are guided by their scouts' evaluation of the long-term potential of the draftees under consideration. While they are sometimes portrayed as adversaries, in Moneyball and elsewhere in the media, scouting and statistics are each crucial tools employed by a team's Baseball Operations staff to understand the value and projection of the players they have, and those they hope to acquire. At the Major League level, we know much more about a player's competition and the ballparks they play in, and we're able to acquire much more data that can help explain why a player's performance might deviate from his true level of skill, so it is often prudent to lend increased influence to statistical analysis when making Major League roster decisions. However, for players in lower leagues, statistics become much less valuable, as the level of competition the player faces can be highly variable and many other factors can make statistics deceptive. For amateur evaluation, teams must rest decisions heavily on scouting, especially because the team's interest is almost entirely in what the player can be in the future rather than his actualized skill set at a particular time. However, I believe that using statistics to study draft outcomes can help us to identify trends that provide context for a scout's evaluation. A scout may be intrigued by a college first baseman, for example. However, the data show that college first basemen average only .215 WAR at the Major League level (Table 2), the worst of any college position and ahead of only the pitiful production of the eight high school second basemen drafted, who put up an average of .038 WAR and avoided totaling zero production solely through Brandon Moss' 0.3 WAR over his first three years in the league. With that knowledge, the team might be wary of selecting that first baseman despite the scout's endorsement, and instead might consider looking elsewhere for another player in whom their scouts have confidence at a position where other draftees have produced better long-term results.

Positional Value

Third basemen have, by far, the highest expected pre-arbitration value among draftees at all positions (Table 3). On average, a third baseman will produce 1.35 wins before reaching arbitration, nearly double the overall average of 0.72. Third base also displays one of the largest gaps between expected value of college and high school draftees, as shown in Table 4, with college third basemen proving far more valuable in this sample. At 1.62 WAR, college third basemen had the best expected value overall of any position/age grouping.

At 1.06 WAR, outfielders produced the second most average value among positions (Table 3). While the outfielders couldn't outpace the production of the third basemen, outfield draftees still produced an average of 0.34 WAR more than any other position. Outfield production was more balanced among the age groups than it was for third base, with high school outfielders producing an average of 1.17 WAR as compared to an average college outfielder's production of 1.01 WAR (Table 4). The mark for high school represents the highest expected value for any position among high schoolers.

The large gap between the third base and outfield groups and the rest of the positions suggests that this could represent a trend worth exploring further. I find it interesting that these positions are located centrally on the defensive spectrum, in terms of their defensive difficulty and the resulting offensive expectations of the position. (James, 1988) Third base and outfield are both less demanding defensively than catcher and the middle infield positions, which require strong defenders and as a result generally feature some of a team's weaker hitters. Likewise, first basemen (and DH) are generally employed for their talent with the bat, while any ability with the glove is essentially a bonus. Near the center of the spectrum, outfielders and third basemen balance glovework with hitting ability. I suspect that this balance is a key factor in the higher

expected value for those positions. I would be interested to see whether this trend extends to the individual players at these positions, in terms of whether we would expect better long-term production from a player who offers some offensive and defensive skills rather than a player of a similar talent level who is expected to produce the large majority of his value in one phase of the game.

These data show a somewhat radical departure from Jazayerli's findings using data from the 1990s. As Jazayerli suggests, the systemic shift in baseball to greater reliance on statistics has made a massive impact on teams' draft strategies, and several college positions which Jazayerli found to be exceptionally fruitful have had a much worse track record since those changes began in earnest in the early 2000s. Jazayerli found that college third basemen produced an expected value fairly similar to the average production of college players overall. Jazayerli also found that college players drafted in the first round produce, on average, roughly 55% more value than their high school counterparts. When teams began to study the draft with a more analytical eye, this inefficiency became a major point of emphasis, and as a result, a larger number of college players were selected in the early rounds. In recent years, high school players have narrowed that performance gap, as is demonstrated in these data. This is especially true at first base, which Jazayerli found to be the most valuable position on average among college players but which for the sample in this study proved the least valuable of all college players.

Interaction of Age and Position

As noted, overall, high school players were more valuable than college players in this sample, producing on average 0.09 WAR in additional value (Table 2). However, when viewing the interaction of age and position, there appears to be a significant advantage for college players at the more demanding positions in the defensive spectrum, while high school hitters at offense-

first positions hold a clear advantage over college hitters. At 0.04 WAR, high school second basemen produce the lowest expected WAR value among any group, with shortstops (0.24) and catchers (0.50) rounding out the bottom three positions by expected value among high school draftees (Table 4). However, college players at those same positions have been able to add value, unlike their less experienced counterparts, with the middle infield positions both producing an expected value of 0.89 WAR, behind only third base and outfield among college positions. Unlike the other defensively challenging positions, college catchers have fared poorly as well, so teams may be better off searching for catchers internationally or on the free agent or trade markets.

Conversely, first base produced the second highest average value among all positions for high school draftees, behind only outfield, at 1.0 WAR. As discussed earlier, first basemen had the lowest average value among college draftees, at only 0.22 WAR (Table 4). This is by far the largest gap in favor of the high school players, so this result is worth investigating to at least attempt to formulate a hypothesis as to why it is so.

These trends may stem from the difference between tools that can be developed through instruction and those that are intrinsically part of a player's game, as well as the difficulty or ease in scouting different types of tools. In the value of college players and relative lack of production of high school players at defensive positions, I believe a major factor is the difficulty of projecting long-term defensive ability. At 17 and 18 years of age, high school players often still have some physical growth in their future. While that growth has a number of positives, it can also act as a negative in reducing the player's athleticism. Because the number of players who can effectively man a middle infield position is so low, these players are highly valuable, and teams will often attempt to let players continue to develop at a middle infield position for as long

as possible. For most, however, their growth and aging eventually dictates that they cannot handle the rigors of playing shortstop, second base, or catching at the Major League level.

When this happens, the player moves to a less demanding defensive position, often shifting from short to third, second to the outfield, or catcher to first base. Once that shift happens, the player encounters a much higher set of offensive expectations, because, on average, other players at his position will produce at a higher level. Even if the player was an above-average hitter as a middle infielder, the move across the defensive spectrum can remove much of the player's value because the same production doesn't fulfill the requirements of the new position. As high school players at demanding defensive positions grow out of their positions, they're left in a somewhat in-between state, as they can neither fulfill the offensive expectations of their new position nor the defensive requirements of their old one.

For college players, these positions produce much more value, likely because scouts can much more accurately determine whether that player has the long-term ability to remain at the defensively demanding position. At this stage, much more of a player's growth is complete, and if his motions on defense continue to be smooth and athletic a scout may determine that he will likely remain outstanding on defense and be able to one day man that position for the Major League club.

For first basemen, the relative value of high schoolers as compared with college players may stem from the importance of power at the position. As a tool, power often develops quite early, and the projection of future power potential is often one of the most enticing skills a prospect can display. As a result, many of the high school prospects with impressive power projection will be identified by teams and drafted and subsequently signed, preventing them from honoring their college commitment and from being drafted as a college player in a future draft.

This leaves a weak resulting college first base class, as is reflected in the data, because most of the players who do decide to honor their college commitment do so because their skill level did not warrant a large draft bonus.

Similarly, high school pitchers fared better than their college counterparts, with the younger athletes producing an average of 0.76 WAR and college pitchers averaging only 0.52 (Table 4). While MLB pitchers employ many strategies as an avenue to success, one of the most basic and attractive skills for any pitching prospect is fastball velocity. While velocity, like power, can develop over time with the addition of strength, it largely relies on the pitcher's body and genetics. Because big arms are hard to find, pitchers with blue-chip velocity aren't usually passed over in the draft, especially as high school arms with positive long-term projection. As a result, many of the players who fall in the draft and do end up arriving on campus are pitchers who rely on command or have a somewhat more developed secondary arsenal, because the ones who light up radar guns receive huge bonus checks and by and large decide to begin their professional careers. Like at first base, the ability to identify early a key trait for the position thins out the college ranks and results in high school draftees proving more valuable in the long run.

Further Investigation

I believe an investigation into the value of balanced vs. offense-reliant players could prove useful, particularly in the evaluation of players at the high-value positions of third and outfield. The success of these positions suggests that balanced players produce more long-term value. My hypothesis is that this would hold on an individual level as well, in that balanced athletes will produce more value, on average, as compared with players who are rated by the

most prominent and respected draft analysts as having a similar skill level on draft day, but who are expected to produce most of their value at the plate.

The study could also be expanded to include players' geographic origins, to further consider the progression over time of James' findings. Draft day discussion often includes the dichotomy between more polished "warm-weather players," those born in warmer parts of the country who have the opportunity to play year round and are therefore more experienced than the "cold-weather" players in their same age group. This can be seen as a positive for the more polished players, but it is also sometimes used to assert that the cold-weather players may have higher potential because their minimal experience leaves the team's player development staff the opportunity to mold and develop the player as they see fit. I would hope to further investigate geographic impact by splitting the draftees into a number of regions, likely consisting of the West, Southwest, South, Northeast, Midwest, Northwest, and Canada. In addition to the warm and cold weather effects, it would be interesting to investigate whether there is any significant long-term difference between players from different baseball hotbeds, such as California, Texas, and Florida.

Finally, more interesting observations might be found in the investigation of the long-term success of these draftees as a function of the teams that draft them. Though this would require an expansion of the data pool to be useful, it could answer a number of interesting questions. At the most basic level, we could learn which teams develop the most valuable draftees overall. Using data from this study, we could investigate which teams are able to most effectively develop players and outpace their prospects' expected value by comparing the expected value of a team's pool of talent to that pool of talent's actual Major League production. Finally, we could determine whether specific teams excel at developing certain types of players

or players at certain positions. This could provide some clues as to which development strategies are the most effective for each type of player, and help determine which organizations' development staffs should be seen as the model for successful player development.

Conclusion

Using data from the 2002-05 MLB drafts, this study isolated a number of trends that may provide insight into how teams can most efficiently develop their draft strategy. Overall, there is no significant difference between the high school and college talent pools without accounting for position. Outfielders and third basemen provided the most expected value among all positions, suggesting possible long-term value in a balanced approach with value added on offense and defense. College players produced more value than high school draftees at demanding defensive positions, while first basemen and pitchers averaged more value when signed out of high school.

While these trends call for more investigation, we can be certain that any sweeping, broad-based statements about the nature of the draft or of evaluating and projecting draftees must be treated with a high degree of skepticism, because of the huge number of variables that are involved in any one projection. In attempting to make simplistic statements, we lose the one value statistics can provide us in amateur evaluation, which is their ability to frame our understanding of a player and place it in the context of the draft as a whole. Amateur scouting will continue to rely heavily on the eyes of experienced, talented evaluators. While many factors in amateur scouting dictate that we must understand the limitations of statistics in the situation, it is important to recognize the value they have as one of a number of lenses through which we can focus our overall understanding of how MLB teams can most efficiently approach the amateur draft.

References:

Jazayerli, Rany. "Doctoring the Numbers: Part 12." Baseball Prospectus, 05 Jun 2006. Web. 11 Dec 2011.

Birnbaum, Phil. "The Value of an MLB Draft Choice." Sabermetric Research, 12 Nov 2007. Web. 11 Dec 2011.

Wang, Victor. "Valuing the Draft (Part 2)." The Hardball Times, 02 Feb 2009. Web. 12 Dec 2011.

Keri, Jonah. *The Extra 2%: How Wall Street Strategies Took A Major League Baseball Team From Worst To First*. New York: Ballantine Books, 2011. 13. Print.

James, Bill. *The Bill James Baseball Abstract 1988*. New York: Ballantine Books, 1988. Print.