Drafting Strategies in the Major League Baseball Draft

Charlie Carluccio
Economics Senior Thesis
Advised by David Owens
April 28, 2011

ABSTRACT:

This study examines how Major League Baseball (MLB) teams have been selecting players in their First-Year Player Draft. The MLB draft is unique because there is a selection issue with two distinct groups of players (high school and college) to select from. An all-encompassing statistic of benefit, WAR, is used to find any distinction in these groups and there is evidence that MLB teams are not drafting according to which group will benefit their teams the most. This study finds that there is value in drafting college hitters within the first 50 picks more than any other player because they become the most successful players in the Major Leagues. In addition, MLB teams will benefit by selecting high school players more often than college players from picks 100-200 because they have the potential to become stars whereas college players are likely to develop into mediocre players.

Introduction

Each year, Major League Baseball (MLB) conducts a prospects draft in June, which consists of 50 rounds, including a supplemental round that occurs in between rounds one and two. The order of each year's draft is determined by the final regular season standings (wins and losses) of the previous season. The draft is done to allocate new players into professional baseball and the draft order is utilized in order to ensure as much parity within the league as possible. A unique aspect of the MLB draft is that, unlike the National Basketball Association and the National Football League, the MLB allows young men to enter the prospects draft once they have completed their high school education. Major League teams, therefore, have the choice of selecting players that have either (i) just completed high school or (ii) completed their Junior year of college (however, college players 21 years or older may be drafted at any time). To be more specific, if a high school player decides to forgo the draft after they have completed high school and enter a four-year institution, that player must wait three more years, or until they turn 21, to be draft eligible again. In other words, MLB teams have the choice of selecting from two distinct groups of players, high school or college. For this reason, one can examine the drafting strategies of MLB teams with regards to these differentiated groups. The final draft eligibility rule is only players that live in the U.S., Canada or Puerto Rico can be drafted, foreign players not living in these places are considered "free agents" who are eligible to be signed by any Major League Team.

One aspect that distinguishes the MLB draft from other major sports, like football and basketball, is the systematic difference in drafting strategies of each team. MLB teams and general managers are likely to select the best available player when it is their

designated drafting slot. This is because not a single player drafted will have an immediate impact on the Major League team that selects them. As the former president and CEO of the Expos, Marlins and currently the Tigers, Dave Dombrowski asserts, "Baseball people refer to the Minor Leagues as the player development phase of their careers. Our primary concern there is to develop players for the Major League level."[1] Therefore, teams are likely to select the players projected to have the most success and greatest development through the Minor Leagues and who will eventually provide their team with the most wins possible.

The choice in the drafting process allows one to investigate the eventual success of the two distinct categories of high school and college prospects along with how Major League teams have been drafting. Teams are conflicted when deciding on which player to draft because of the dissimilar advantages each player, high school and college, will bring to their team. Orioles scouting director Joe Jordan explains, "Sometimes with high school players we just feel like if we can get them in our system, we can develop them the way we want…But there is also the college player that may be more of a proven guy. You've had more looks and more of a track record with that player."[2] But does this mean that Major League teams are weighing these issues appropriately? Or is there a difference in which group of players teams are drafting, especially in the early rounds of the draft where teams expect to acquire elite players who they believe have the highest likelihood of not only making the Major Leagues, but also benefiting their respective team in terms of adding the most wins possible.

---

[1] Go Pro Baseball Wise: Minor League Baseball, http://www.baseballwise.com/club/minorleague.html (2008)

[2] Steve Melewski, *Jim Callis and Joe Jordan on high school vs. college in the draft,* http://www.masnsports.com/steve_melewski/2011/03/jim-callis-and-joe-jordan-on-high-_school-vs-college-in-the-draft.html (March 2011)

Table 1: MLB Draft: Selection Breakdown (1980-2005)

Picks	College	High School	College Hitters	High School Hitters	College Pitchers	High School Pitchers
1-50	675	611	294	350	381	260
51-100	651	639	300	373	351	266
101-150	752	544	367	298	385	246
151-200	807	485	383	270	424	215

Table 1 shows the distribution of how MLB teams have drafted from the years 1980-2005. As one can see, the number of college and high school players selected in the first 100 picks is very similar, college players being drafted at a slightly higher rate. From picks 100-200, however, teams have been drafting many more college players than high school players. One reason this could be the case, though, is because in this pick range more high school players are choosing to attend four-year institutions instead of entering into the Minor Leagues.

Breaking down the data even further, table 1 displays the breakdown of how MLB teams have drafted college hitters, high school hitters, college pitchers and high school pitchers. In this split, it is clear that teams select college pitchers the most often (leading each split except picks 51-100), followed by college hitters and high school hitters with

high school pitchers being selected the least. This paper aims to either (i) agree with how teams have been drafting, or (ii) find a fault in the decision-making that has been going on in the MLB draft since 1980.

WAR

In order to judge the "value" of players, this paper will use the sabermetric statistic entitled WAR, or wins over replacement. WAR is a statistic used in baseball that informs one on how many additional wins a particular player gives his team as opposed to a bench or minor league player. To be more specific, WAR is a metric that combines a player's offensive and defensive contributions on the field and then compares him to a replacement level player for that position, with replacement level considered the level at which a player can be easily replaced by a bench or minor league player.[3]

Additionally, it is important to note the nature of WAR as a statistic. It is a cardinal measure of value. For these reasons, WAR of drafted players from 1980-2005 will be used to help determine whether or not MLB teams have been utilizing the most beneficial (in terms of wins, not monetary value) drafting strategies.

[3] See appendix for a more in depth explanation of how WAR is calculated

Literature Review

Victor Wang (2009) conducts a study on *Valuing the Draft* (Part 2), where he attempts to value the MLB draft in terms of high school and college players. More specifically, Wang aims to discern any differences in the success of high school and college players in the Major Leagues. In the study Wang includes first round picks, supplemental/second round picks and third round picks from 1990-1997. Wang then divides the picks into high school hitters, college hitters, high school pitchers and college pitchers. He separates his results based on the rounds and finds that one can better predict the success of both high school and college hitters, compared to high school and college pitchers, using WSAB, win share above bench player (a statistic very similar to WAR). This translates to a finding that saw the average WSAB for all hitters, high school and college, to be higher than WSAB for all pitchers, high school and college.

Aside from this finding, Wang does not provide any additional results that can differentiate between high school and college players. However, Wang does conclude that because of his findings from the first round, MLB teams should be drafting hitters ahead of pitchers because the probability of their success in the first round is greater.

Jacob M. Lee conducts two similar studies that aim to find any difference in the eventual success of high school and college players. His two studies separate pitchers and hitters and his data only looks at the top 10 draft picks in the first round from 1980-1999. He implements a grading scale (that he has defined himself), instead of a statistic like WAR, to compare high school and college players. His grading scale is much more informal than WAR and the "grades" he assigns can be debated, while WAR is more formal because of its cardinality and strict formulation. However, the grading scale leads

Lee to conclude there is no evidence to suggest that either college or high school hitters should be preferred over the other, but does find evidence to suggest that college pitchers should be selected ahead of high school pitchers in the first 10 picks of the draft. While this thesis does not directly relate to Lee's paper, it did motivate the idea of investigating any differences between high school and college players in terms of bins of picks.

Data Analysis

The data for this project, which is compiled from baseball-reference.com, includes all players selected in the first eight rounds of the MLB draft from 1980-2005. By expanding the data to eight rounds one is able to study in greater detail the way in which decisions are being made in the MLB draft and whether or not these decisions have an inherent flaw to them. The hypothesis of this project is that Major League teams and general managers have been conducting incorrect drafting strategies by electing to choose high school and college players at about the same rate within the first 100 picks because the expectation of success for college players, hitters and pitchers, can be better predicted than the success of high school hitters and pitchers. In addition, this paper hypothesizes that MLB teams are also using incorrect drafting strategies in selecting many more college players from picks 101-200 because high school players have more of a chance to become stars as opposed to mediocre players.

The data contains many players who never made it to the Major Leagues and the WAR for these players has been recorded as zero. While the WAR of some players is negative, the lowest being -5.1, zero is the appropriate number to give these players who never made it to the Major Leagues because their teams recognized they would not be successful Major League players. To omit WAR for players that never made it would not

provide accurate evidence for or against the current drafting strategies and decision making by MLB teams because any results would only suggest revisions in strategies based on players who will make it to the Major Leagues. Unfortunately, many players do not make it to the Major leagues, and thus, these players must be accounted for in the data.

The data consists of 6,012 players, 2,560 of which were drafted out of high school and 3,452 were drafted out of college. Again, the data is limited to the first eight rounds of the draft and this is the summary of those first eight rounds. As stated above in the paper, college players are taken at a slightly higher frequency than high school players in the first 100 picks, but many more college players are selected from picks 101-200. As noted earlier, this could be the case because high school players are more likely to attend college than enter into the Minor Leagues, or it could expose more about how general managers draft. There could be a selection issue present and with less known about high school or college players, general managers may be more likely to select from one of the aforementioned groups (college or high school players).

Figure 1: Hitter and Pitcher Averaged WAR Per 5 Picks on Pick Number

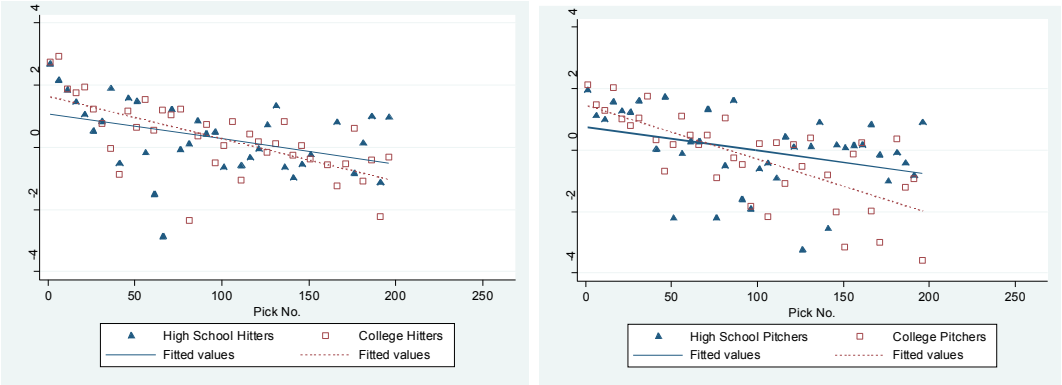

Figure 1 displays the graphs of all hitters (left) and pitchers (right), separating high school and college, and their career WAR's. The WAR's do not contain every

single player, rather the points represent the average WAR of every five picks (the first point is the average WAR of picks 1-5 and point two is the average WAR of picks 6-10, etc…). The graph takes the players who did not make it to the Major Leagues into account with their WAR's entered as zero. Further, table 2 displays descriptive statistics of these graphs, and reveals additional information about the data.

Focusing on the graph of all hitters, left, one observes clear evidence that college hitters selected within the first 75-90 picks of the draft become stars (have consistently high WAR's) in the Major Leagues more often than high school hitters drafted within the first 75-90 picks. After pick 90, the probability in predicting the success of college hitters appears as if it disappears. Table 2 exhibits a larger variance for high school hitters from picks 101-200, 1.212, than college hitters, 0.436. This is not surprising, however, because the observations of the college hitters are more centralized around its best-fit line than for high school hitters.

The graph of the pitchers, right, appears to show a slight edge, in terms of WAR, for college pitchers within the first 75 picks, but it is not as great as the difference is for college and high school hitters. Additionally, table 2 displays larger variance for high school pitchers from picks 51-100 and 101-200, 2.89 and 0.511, compared to college pitchers in these pick ranges, 0.935 and 0.312).

Table 2: Descriptive Statistics

	Mean	Std. Dev. (Variance)	Min	Max
College Hitters Picks 1-50	6.382	6.12 (37.454)	.426	18.59
HS Hitters Picks 1-50	5.302	4.124 (17.007)	.595	14.656
College Pitchers Picks 1-50	3.957	2.601 (6.765)	.512	8.407
HS Pitchers Picks 1-50	3.634	2.128 (4.528)	-.068	7.056
College Hitters Picks 51-100	2.205	1.360 (1.85)	.096	4.612
HS Hitters Picks 51-100	1.644	1.376 (1.893)	.057	4.386
College Pitchers Picks 51-100	1.354	.967 (.935)	.163	3.039
HS Pitchers Picks 51-100	1.35	1.7 (2.89)	.11	5.023
College Hitters Picks 101-200	.919	.66 (.436)	-.03	2.333
HS Hitters Picks 101-200	1.004	1.101 (1.212)	-.153	3.796
College Pitchers Picks 101-200	.564	.559 (.312)	-.108	1.511
HS Pitchers Picks 101-200	1.028	.715 (.511)	.039	2.467

Regression analysis has been run on four different dependent variables. These four dependent variables are the natural log of college hitters career WAR, the natural log of high school hitters career WAR, the natural log of college pitchers career WAR and the natural log of high school pitchers career WAR (all in the bins of five picks). The independent variable for each of these dependent variables is pick number. The analysis aims to figure out the percentage drop in career WAR as the pick number in the draft

increases. In addition, the standard errors and R^2 values reveal vital information necessary to help form conclusions.

Table 3: Separated WAR's and Pick Number

	College Hitters	High School Hitters	College Pitchers	High School Pitchers
Pick Number	-.014*** (.000) s.e.=.00239 R^2 = .4681	-.008** (.014) s.e.=.00308 R^2 = .1665	-.018*** (.000) s.e.=.00278 R^2 = .5253	-.0077** (.027) s.e.=.0034 R^2 = .1250

p-value in parenthesis
*indicates statistical significance at 90% confidence
**indicates statistical significance at 95% confidence
***indicates statistical significance at 99% confidence

Table 3 displays the results from the regressions run. The log of the dependent variables is used to create a linear relationship with pick number. Both college hitter and pitcher coefficients are significantly different from zero at the 99% confidence level with t-values of -5.71 and -5.98, respectively. Each high school hitter and pitcher coefficients are significantly different from zero at the 95% confidence level with t-values of -2.61 and -2.30, respectively. These results help conclude there is a percentage drop in WAR as pick number increases. This drop occurs at a higher percentage for college hitters and pitchers in comparison to high school hitters and pitchers. This coincides with the findings from the previous graphs, which display the elevated predictability of success for both college hitters and pitchers. Even more important, however, are the standard errors and R^2 values of the dependent variables. As one can perceive in table 3, the standard errors are lower for college players and the R^2 values are higher for college players. This conveys the predictability of the regression equations, which assert the

increased accuracy of the equations for college players and thus, higher variance of high school players.

Discussion

In examining the graphs and descriptive statistics again, it is evident that the variance of high school hitters is greater than for college hitters. Going back to the numbers stated earlier, more college players are selected from picks 101-200, the picks that show the increased variance in high school hitters, which leads to a discussion that it may be worth selecting a high school hitter in this pick range rather than a college hitter because there is more that Major League teams can gain. MLB teams will have more to gain because, while the accuracy of predicting the success of college hitters does appear to be greater, none can be predicted to become grand contributors to their teams, whereas some high school hitters can be predicted to achieve a substantial WAR. In essence, the college players selected in the 101-200 pick range will achieve a WAR of a mediocre player whereas a high school player has the potential to become a star because of their higher variance.

Also, the variance of high school pitchers from picks 101-200 is more than college pitchers, but again, not as great as the difference for hitters. WAR is an attempt at an all-encompassing statistic valuing the benefit across all positions, but in thinking strictly about pitchers, it may also benefit general managers to select more high school pitchers in this pick range because they have a greater probability of achieving more substantial WAR's than those of college pitchers. Again, college pitchers drafted from picks 101-200 will become mediocre players but high school players in this range have the potential to become major contributors for their Major League teams.

Figure 1 and table 2 also allow for a comparison between hitters and pitchers and it can be argued that a sound strategy for general managers of Major League teams to draft hitters ahead of pitchers. The mean WAR's for all hitters, especially college hitters, in picks 1-50 is greater than those of all pitchers and therefore Major League teams, assuming their objective is to draft players who will provide their teams with the most wins possible, should select college hitters followed by high school hitters before picking any pitchers. This finding agrees with Wang in that he also advises that drafting hitters ahead of pitchers is a sound strategy for MLB teams to employ.

Conclusions

MLB teams do seem to be conducting incorrect drafting strategies. Within the first 75 picks of the draft, there is value in selecting college hitters the most frequently because the college hitters expected to become stars do become stars compared to high school hitters and all pitchers. Currently, teams are drafting both high school hitters and college pitchers at a higher frequency, thus decreasing their expectation of success, in terms of wins, with that selected player.

Further, it is in the MLB teams best interest to draft high school players from picks 100-200 because of their higher variance. At present, MLB teams are selecting college players in this pick range at a much greater rate than high school players and thus are giving their team less of a chance at adding substantial wins with these college players because the predictability of their mediocrity in the Major Leagues is inevitable compared to the high school player who has the potential to achieve greatness. Senior editor for *Baseball America* agrees with this conclusion stating, "more college guys as a

percentage get to the majors, but the difference is almost all fringe, cup of coffee guys. More high school guys, as a percentage, become stars or superstars."[4]

These conclusions are important because they stand to correct the misguided drafting strategy that MLB teams have implemented since at least 1980. This information suggests changes that give MLB teams greater chance at success in the future with the players they are drafting. The first 200 picks is about 8 rounds of the draft, consequently teams will only have around 8 picks in this range. Of these top 8 picks, it is instrumental for teams to select players that have the greatest probability of adding the most wins possible to their teams and this has not been occurring. There is value in MLB teams doing away with their current drafting strategies within the first 200 picks and attempt the conclusions found in this paper in order to give themselves the best chance at adding more wins per season in the future.

Future Work

This project has looked solely into WAR, and not at the signing bonuses of different players. I am currently in the process of putting together a separate data set that includes the signing bonus information for every player. Unfortunately, *Baseball America* does not go back to 1980, but the years and players are arbitrary. This will be a new data set that includes all players drafted in the first 8 rounds from the years 2003-2010. This data will not tell anything about benefit (WAR), but monetary value (in terms of signing bonuses when a player signs a contract after the draft). By separating high

---

[4] Steve Melewski, *Jim Callis and Joe Jordan on high school vs. college in the draft,* http://www.masnsports.com/steve_melewski/2011/03/jim-callis-and-joe-jordan-on-high-school-vs-college-in-the-draft.html (March 2011)

school and college hitters and pitchers again, any glaring differences in value will present itself. It may be the case that college hitters are the most expensive and all pitchers are much less expensive and that is why teams still pick pitchers at about the same rate in the first four rounds as hitters even though hitters benefit their teams much more.

References

Carroll, Will. *Labrum, It Nearly Killed Him*, http://www.slate.com/id/2100895/

Forman, Sean. *Player Wins Above Replacement*, Retrieved from February 7, 2011 from http://www.baseball-reference.com/blog/archives/6063

Lee, Jacob M. *MLB Draft: Should You Draft College or High School Position Players*. Retrieved from February 3, 2011 from http://www.associatedcontent.com/article/5514838/mlb_draft_should_you_draft_college.html?cat=14

Lee, Jacob M. *MLB Draft: Should You Draft High School or College Pitchers.* Retrieved from February 3, 2011 from http://www.associatedcontent.com/article/5483930/mlb_draft_should_you_draft_high_school.html?cat=9

Melewski, Steve. *Jim Callis and Joe Jordan on high school vs. college in the draft,* http://www.masnsports.com/steve_melewski/2011/03/jim-callis-and-joe-jordan-on-high- school-vs-college-in-the-draft.html

Wang, Victor. The Hardball Times. *Valuing the Draft (Part 2),* Retrieved from February 3, 2011 from http://www.hardballtimes.com/main/article/valuing-the-draft-part-2/

Appendix

Baseball Prospectus calculates WAR for hitters by looking at a number of run statistics that culminate in RAR, the sum of all run components. The components include:

--R***bat***, batting runs are a linear weights formula utilizing custom weights depending on team runs scored and run scoring environment. This essentially looks at the team and run scoring environment a player is dealing with. This will takes into account a player who scores and produces runs on a team that is not very good offensively more so than for a player with similar statistics who plays for a better offensive team.

--R***bsr***, Baserunning events like stolen bases, advancing on passed balls, going first to third on a single. For 2010, this is currently just SB and CS data.

--R***roe*** gives a bonus or penalty depending on whether the player reached on error more often than average. Typically +/- 5 runs at the most. This is a relatively minor skill, but does in fact appear to be a skill. A player who reaches on error more often usually hits balls very hard into play and this will help their WAR.

--R***dp*** gives a bonus or penalty depending on whether the player grounds into more or fewer double plays than expected. Typically +/- 5 runs at the most. The more double plays a player hits into the more the negative impact will be on their WAR.

--R***field*** is the value in runs of all aspects of the player's fielding (fielding balls, outfield arms, turning the double play, controlling the running game). Good fielders are rewarded and below average fielders are penalized.

--R***pos*** is the value in runs of playing a particular position. This is inversely proportional to the hitting ability of the average player at that position. 2009 values per 1,350 innings (150 games): Catcher 10 runs, Shortstop 7.5 runs, Second Base 3 runs, Third Base 2 runs,

Center Field -2.5 runs, Right and Left Field -7.5 runs, First Base -10 runs, Designated Hitter -15 runs. This is an attempt to evenly match all positions to truly have WAR be all-encompassing across all positions.

--**R*rep*** is the value of an average player over a replacement player given the player's playing time. Replacement level is set at around a .320 team W-L percentage. AL's is 22 runs per 650 PA and NL's is 18 runs per 650 PA. A player's PA is the smaller of actual PA and 4PA/G*G in order to not overvalue leadoff hitters.

--**RAR** is the sum of all of the run components.

--**WAR** converts runs above replacement to wins above replacement. Runs per win is dependent on the run scoring environment the team played in. Currently using 10 runs/win for 2010 data.[5] For pitchers, the calculation is a bit different:

--**R** pitcher's actual runs allowed

--**R*rep*** runs allowed by a replacement level pitcher given this pitcher's defense, park, strength of opponents, and role (replacement level is different for starters and relievers). The number presented includes the defensive component, R*def*, already.

--**R*def*** runs above or below average for this pitcher's team defense. The team's overall defense weighted by the percent of balls in play allowed by this pitcher. A pitcher is dependent on the defense behind them, so it makes sense to adjust WAR accordingly.

--**aLI** is the average leverage index this pitcher pitched in. To convert RAR to WAR, the pitcher's runs above replacement is weighted by the average of their leverage and 1.00 (due to bullpen chaining). Average leverage is 1.00. Above 1.00 includes high leverage and below 1.00 indicates lower leverage.

---

[5] Sean Forman, *Player Wins Above Replacement*, http://www.baseball-reference.com/blog/archives/6063

--**RAR** runs saved above replacement. This is not weighted by leverage yet.

--**WAR** RAR/(Runs per win), Runs per win varies with the run scoring environment of the pitcher and how that pitcher changes the environment. Typically around 10 runs per win.