Identifying Optimal Draft Strategies in a Baseball Simulator

Hubert Luo [1]

Abstract

Scraping and analysis of 108,488 draft picks from 23 seasons based on various draft metrics in a baseball simulator.

1. Introduction

BrokenBat is an online baseball simulator created by Steve Muller of Solana Beach Design in 2009. Out of a total of 756 teams on BrokenBat, there are more than 600 active users who each manage a baseball team consisting of fictional, computer-generated players. (Muller, 2020)

Players are initially distributed to teams through the use of a draft. Each week, teams receive a randomly generated list of ten prospects from which they can draft from, with control over which draft source pool and position they wish to draft from. The overall pool of prospects is shared and replenished with newly generated players only at the beginning of each season. There are eight rounds in the present-day draft.

Prior to May 2019 before BrokenBat season 2041, there were ten rounds in the draft instead of eight rounds today and only five prospects from which a manager could draft from instead of the ten options in the new system. (Muller, 2019) In addition, a draft prospect's potential was hidden while selecting the prospect so under this old draft type, managers selected prospects to draft solely on their scouting report. Player attributes, statistics, and potential were all previously hidden from the manager prior to making their selection. After May 2019 and starting in BrokenBat season 2041, a new draft mechanism was introduced. Player attributes, statistics, and potential are all visible to the manager prior to making their selection.

This paper aims to investigate the historical efficacy of drafting based on a player's potential based on the draft round, year, source pool, and player position. This would be impactful as it allows for more evidence-based decision-making in selecting a draft pool or player position from which to draft players from. It also gives managers a more nuanced view of the draft process with more realistic expectations. The analysis also aims to examine the changes as a result of moving from the old draft to the new draft.

[*]Equal contribution [1]Department of Statistics, University of California, Berkeley.

Copyright 2020 by the authors.

2. Approach

2.1. Previous Work

This paper builds on a previous paper under the old draft system that focused on developing models for predicting potential based on scouting report, age, and draft pool. (Luo, 2018) This was important previously under the old draft system as potential was previously not revealed to managers until after the draft selection was made. In the new draft system, however, there have been shifts in drafting patterns now that potential is known from the outset.

2.2. Data Scraping

From each of the 756 teams in BrokenBat, a total of 108,488 draft picks were analyzed. These spanned 23 seasons, from 2023 to 2045. This data was scraped from each team's draft page, which contains information about the draft selection, source, and name, in addition to current age, potential, assignment, and skill index (SI), a metric that measures the overall attributes of a player. For players drafted after the first round of the 2043 draft, i.e., from the second round of 2043 onward, a player's draft potential and draft SI are also displayed. This allows for data to be collected on how a player's potential has changed from the time they were drafted to the day the data was collected, 23 June 2020.

Permission was received from the game's administrators prior to the first round of scraping was performed in 2018, and a ten second delay between scraping draft information from a team was instituted to reduce the server load.

2.3. Data Description

Out of the 108, 488 draft picks, information was gathered on their draft source, position, draft/current age, name, current assignment, draft team/team ID, draft year, draft/current SI, and draft/current potential. In addition, based on the draft year, a variable corresponding to which draft system that player was drafted under was created.

Table 1. Average current and draft potential of all players under the old and new draft systems by draft round.

RD	AVG CUR POT (OLD DRAFT)	AVG CUR POT (NEW DRAFT)	AVG DRAFT POT (NEW DRAFT)
1	12.623	12.914	12.972
2	12.366	12.675	12.695
3	12.146	12.481	12.376
4	11.943	12.187	12.009
5	11.712	11.998	11.729
6	11.510	11.757	11.474
7	11.279	11.657	11.297
8	11.100	11.495	11.135
9	10.876	NAN	NAN
10	10.643	NAN	NAN

A player's potential is a number between 1 and 20 that determines roughly how many total SI points a player can have, with some random variance. Under the old draft system, potential was a fixed number that couldn't change. Under the new draft system, however, potential is a number that may change over time.

The average age of draftees was 19.101, and the average draft SI was 62.486. The average current potential of draftees was 11.776, while their average potential when drafted was 12.035. Overall, 30.436% of draftees had at least 13 potential currently or when drafted, while 10.292% of draftees had at least 14 potential currently or when drafted.

3. Results

3.1. Draft Round

A player's potential is highest for the earlier rounds and decreases as the rounds go by, as seen in Table 1 which displays the average current and draft potentials of various players. Note that potential did not change under the old draft system, so therefore the current and draft potentials are equal under the old draft system.

Furthermore, the average current and draft potentials are both higher under the new draft system than under the old draft system for almost all rounds. This is especially noticeable for the earlier rounds, when the averages under the new draft system are higher by about 0.3. The new draft only consists of eight rounds while the old draft consists of ten rounds, so hence there are clearly no averages available for the now-outdated rounds 9 and 10 that existed under the old draft.

Figure 1 displays the distribution of players' current potential by draft round and draft type. Again, generally the pattern of higher draft potential in the earlier rounds and lower draft potential in the later rounds are observable. The first few rounds consist of primarily prospects with 12, 13, or 14 potential, before only primarily consisting of prospects with 10, 11, or 12 potential by the end of the season for the later draft rounds.

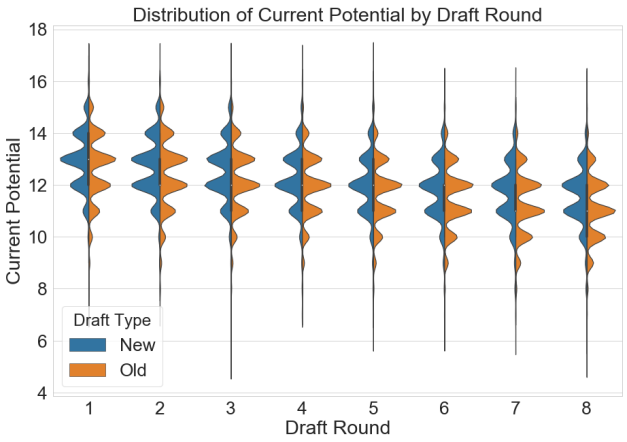

Figure 1. Distribution of current potential by draft round under the old and new draft types.

Under the new draft system, it is again noticeable that generally there are more prospects with a higher potential than there were under the old draft system, as seen in Figure 2. At the same time, however, as seen in Figure 1 there are also more with a much low draft potential, for example there are more players with 11 potential in the first few rounds under the new draft system than the old. This can be explained by the "weeding" that some managers conduct, where they purposely select and then release a player with very low potential.

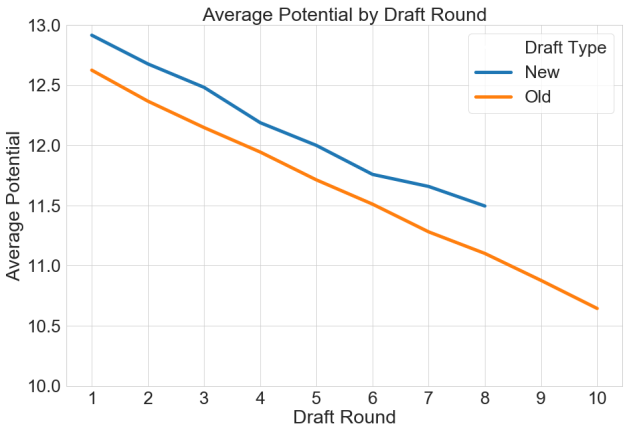

Figure 2. Current average potential by draft round under the old and new draft types.

Table 2. Average proportion of draftees with at least 13 potential by round under both the old and new draft systems by draft round.

ROUND	AVG PROP 13+ POT (OLD DRAFT)	AVG PROP 13+ POT (NEW DRAFT)
1	0.564	0.690
2	0.475	0.597
3	0.395	0.528
4	0.319	0.421
5	0.247	0.346
6	0.195	0.267
7	0.149	0.241
8	0.121	0.181
9	0.083	NAN
10	0.055	NAN

This change from the old draft system to the new draft system is perhaps most visibly noticed in Table 2, which displays the average proportion of draftees with at least 13 potential by round under both the old and new draft systems by draft round. Under the new draft system, there is a 22.34% percentage increase from 0.564 to 0.690 in the likelihood of drafting a prospect with at least 13 potential in the first round.

3.2. Draft Year

Under the old draft system, there are only minor deviations in the current potential of draft prospects, as demonstrated by the years prior to 2041 in Figure 3 which displays the distribution of current potential in the last ten years. The years prior to 2041 under the old draft system all clearly have a roughly similar distribution of current potential.

However, this repetitiveness is no longer true starting in 2041 under the new draft system. As a result of the creation of this new system, 2041 saw a noticeably higher proportion of players who had higher potential be drafted which corresponds with anecdotal evidence obtained from multiple managers, where 2041 is often referred to as the "super-draft" year as a result of the higher-than-usual potential attributes found for prospects.

This effect diminishes as the draft year increases, however, where Figure 3 and Table 3 demonstrate that after the first year under the implementation of the new draft system, the current potential of draftees slowly shifts back towards the previous draft distribution. The specific patterns observed include fewer players with 13 or higher potential and more players with 12 or lower potential and these shifts accentuate as the season moves further away from 2041. In particular, the distribution of current potentials of draftees in 2045 appears to be relatively similar to those under the old draft system prior to 2041.

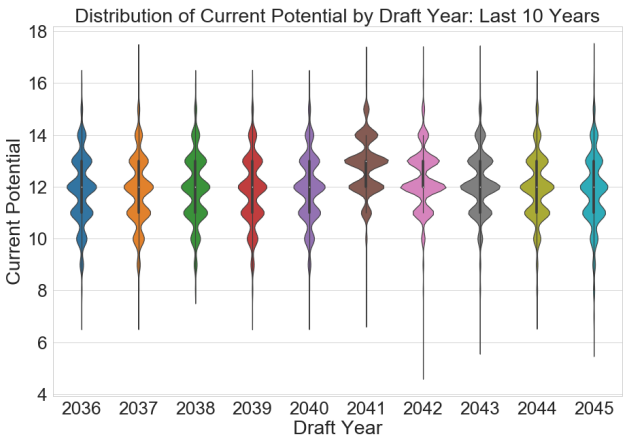

Figure 3. Distribution of current potential by draft year in the last 10 seasons from 2036 to 2045 inclusive.

Table 3. Top 5 draft years by average potential.

DRAFT YEAR	AVG POTENTIAL	SD POTENTIAL
2041	12.777	1.086
2042	12.265	1.119
2043	12.141	1.201
2027	11.986	1.309
2044	11.968	1.279

The 2023 draft only includes data after the third round, so the 2023 draft was excluded from this section in the analysis. Also excluded were all draftees after the eighth round, since those draftees were present only in the old draft and not the new draft, which only has eight rounds instead of ten.

3.3. Draft Source Pool

There are six different possible draft pools from which prospects can be selected from: the Asian Posting System, College, High School, International, Junior College (JC), and the Latin American Academy. Note that JC and College used to be considered together as just College under the old draft system. Figure 4 shows that under the old draft system, the majority of draftees were from the High School, Latin Academy, and College draft pools in that order, with those three draft pools constituting 89.9% of all drafted players under the old draft system.

After the new draft system was instituted however, there is a significant shift from where draftees are picked. The College (including JC) and Latin Academy draft pools now contain the bulk of draftees, combining for 77.9% of all drafted players under the new draft system. There is a patent uptick in the utilization of the college draft pool in particular, and

a noticeable drop in that of the high school draft pool.

Note that players drafted in 2023 did not have the draft pool saved in the team's draft page that was scraped, so all players drafted in 2023 were dropped from this portion of the analysis.

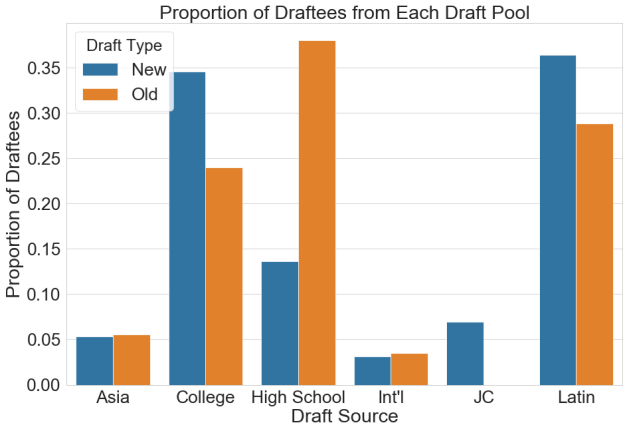

Figure 4. Proportion of draftees from each draft pool by draft type.

Figure 5 demonstrates again that for the most part the new draft system has led to inflated player potential regardless of draft pool and draft round. However, there are instances where the average player potential of a player drafted under the new draft system is lower than it was under the old draft system, specifically in rounds 5-7 of the Latin Academy draft pool for example. In conjunction with the uptick in draftees from the Latin Academy pool as seen in Figure 4, this supports the fact that there is over-selection from managers in the Latin Academy pool and it is diluting the amount of talent available in the later rounds from a potential perspective.

Figure 6 examines in more depth the average current potential of draftees from each draft pool under the new draft system. Average potential remains highest for high school draftees in the first round, similar to results found in previous work for the old draft system. (Luo, 2018) This is followed by College, JC, and Latin Academy in that order.

Deviations from the results in the old draft system, however, are observable starting in the second round when the highest average potential is for junior college draftees, not high school players as it was previously.

Previous work also found that the Latin Academy pool had the highest draft potential on average after the first two rounds. (Luo, 2018) However, Figure 6 demonstrates that under the new draft system, this is no longer the case. College/JC draft picks generally have the highest average potential from the third round onwards, with the inter-mixing of the two pools while drafting especially removes the need

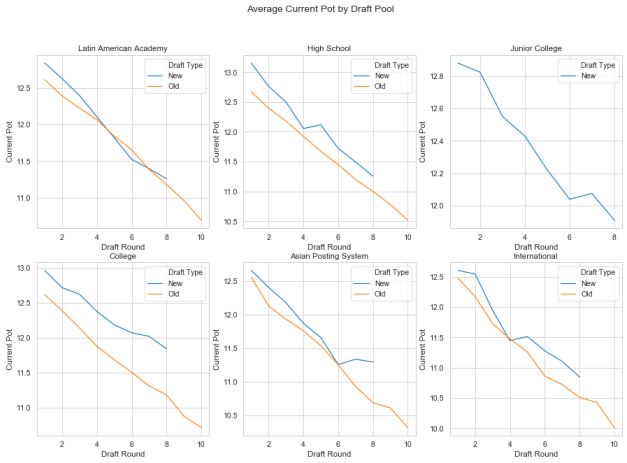

Figure 5. Average current potential of draftees from each draft pool by draft type.

to differentiate between their efficacy from a practical manager perspective. This is followed generally by High School draftees, and then Latin Academy signings. For almost all rounds, players signed through the Asian Posting System and internationally have the lowest average potential.

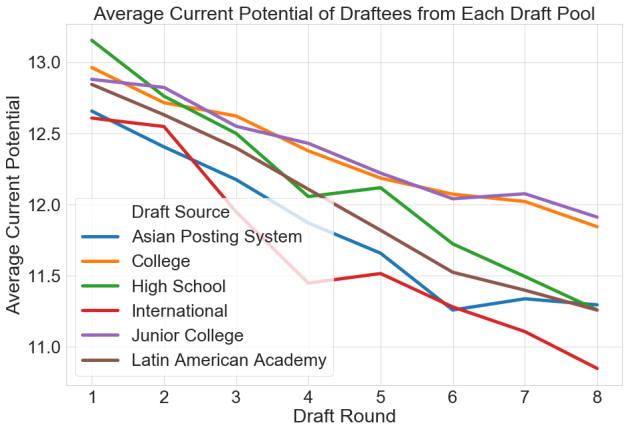

Figure 6. Average current potential of draftees by draft round from each draft pool under the new draft system.

Similar patterns are observed in Figure 7 where there is a palpable increase in the proportion of draftees with at least 13 potential under the new draft system compared to under the old draft system for almost all draft source pools.

Figure 8 demonstrates that the draft pools with the highest proportion of draftees with at least 13 potential is the high school draft pool for the first round, followed by college/JC draft pools for all remaining rounds. Again, these two draft pools the inter-mixing of the two pools while drafting es-

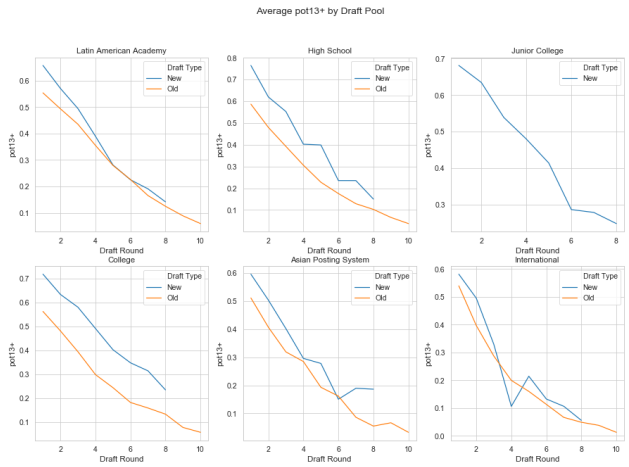

Figure 7. Proportion of draftees with 13 or higher draft/current potential from each draft pool by draft type.

pecially removes the need to differentiate between their efficacy from a practical manager perspective.

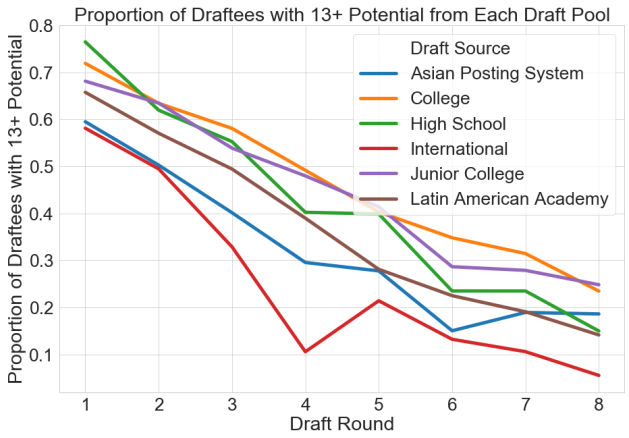

Figure 8. Proportion of draftees with 13 or higher draft/current potential by draft round from each draft pool under the new draft system.

Under the new draft system, draft SI was stored that contained information about the starting SI attribute of each draft pick. This draft SI data is displayed in Figure 9. Players signed through the Asian Posting System have the highest draft SI out of all the draft pools for each round by a significant margin. This follows established knowledge that players signed from the Asian Posting System are generally older and more ready to make an immediate impact on major-league teams with a draft SI average above 70 for each round.

This is followed by college players, who have the second-highest average draft SI out of all the draft pools for each draft round, with average draft SI ranging between 65 and 70 depending on draft round. International, Junior College, and Latin Academy players are interchangeably have the next highest average draft SI. High school players are the most raw and have the lowest average draft SI out of all the draft pools for each draft round, with average draft SI typically ranging between 50 and 60.

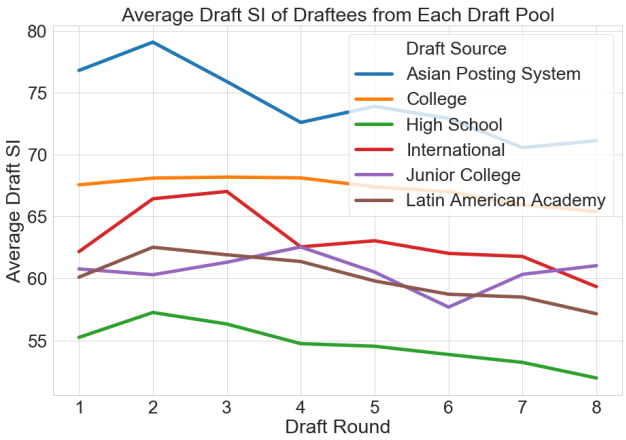

Figure 9. Average draft Skill Index (SI) by draft round from each draft pool under the new draft system.

3.4. Position

Table 4 demonstrates that middle infielders (MIF) such as shortstops and second base have the highest average current potential, followed by outfielders, corner infielders (CIF), catchers, and then pitchers. At each position, higher draft potentials are seen under the new draft system than under the old draft system. Note that Table 4 excludes draftees after the eighth round in order to more accurately compare across the two different draft systems, as the additional ninth and tenth rounds under the old draft would skew the average round and average potential otherwise.

The average draft round of prospects shown in Table 4 suggests that pitchers are being selected relatively early especially, with the lowest average draft round of 3.963 under the new draft system. This is followed by middle infielders, catchers, corner infielders, and finally outfielders for the new draft system. Compared to the old draft system, corner infielders are being selected relatively later while middle infielders are being selected earlier in the draft.

The proportion of draftees at each aggregated position by draft round in Figure 10 demonstrate that a large number of pitchers are taken in the first round especially, with pitchers constituting over 50% of all draftees in the first round under the new draft system.

Table 4. Average draft round and current potential by aggregate position and draft type. Note that MIF corresponds to SS and 2B, while CIF corresponds to 1B and 3B.

DRAFT	AGG POSITION	ROUND	CURRENT POT
NEW	MIF	4.173	12.392
NEW	OF	4.404	12.317
NEW	CIF	4.366	12.269
NEW	C	4.300	12.248
OLD	MIF	4.455	12.121
OLD	OF	4.437	12.102
NEW	P	3.963	12.046
OLD	CIF	4.409	11.997
OLD	C	4.643	11.954
OLD	P	4.216	11.593

In addition, pitchers remain the largest proportion of draftees for all draft rounds, as would be expected since the pitching pool naturally contains more prospects than any of the other individual position player pools. It is highest in the early rounds, and falls quickly by the third round to around 35-40% of draftees.

Outfielders comprise the second-largest proportion of draftees, and the proportion of outfielders drafted at each round increases as the draft progresses. This is then followed by middle and corner infielders, and then catchers. The average proportion of draftees from these three categories stay relatively consistent between draft rounds and are lower than the proportion of outfielders and pitchers drafted for all rounds.

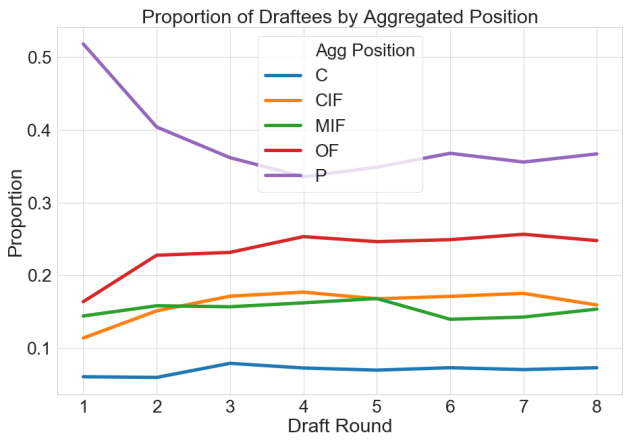

Figure 10. Proportion of draftees by round from each aggregated position under the new draft system.

Figure 11 shows that middle infielders have the highest average proportion of draftees with at least 13 potential in the first round. For the second and third rounds, middle infielders and outfielders are roughly equally likely to have draftees with at least 13 potential. For all remaining rounds, the four aggregate batter positions are roughly equally likely by this metric. Pitchers for all rounds except the first have the lowest average proportion of draftees with at least 13 potential.

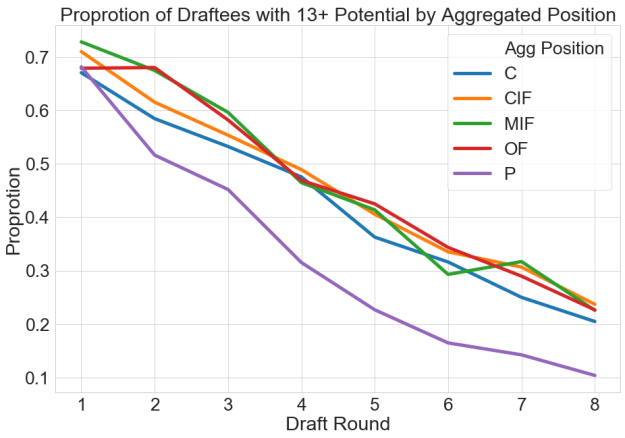

Figure 11. Average proportion of draftees with at least 13 draft/current potential by round from each aggregated position under the new draft system.

3.5. Team

Teams vary drastically in drafting patterns, with a surprising amount of variation in the amount and quality of draft picks. This is demonstrated in Tables 5 and 6 which have the top and bottom five teams by the percentage of draftees with 13 or higher potential, separated by draft type.

Note that for Tables 5 and 6, only teams which drafted at least 90% of the time in either draft type as shown in the frequency column were kept in order to remove some teams which draft very rarely.

The teams with the lowest percentage of draftees who have 13 or higher current potential routinely have fewer than 20% of draftees which fall under this category. On the other hand, some teams appear to consistently draft players with over 60% success at getting a player with at least 13 potential.

Table 6 demonstrates that the top five teams by percentage of draftees with at least 13 potential are all achieved under the new draft system. This can be partially explained by the information asymmetry between the old and new draft.

Since all attribute information including potential is known prior to the draft selection takes place under the new draft system, it is likely that the teams achieving a surprisingly high rate of draftees over 13 potential are using potential as the primary basis from which to draft players instead of

Table 5. Bottom 5 teams by percentage of draftees with at least 13 draft/current potential by draft type. Draft frequency is the percentage of times that team actually drafted a player under a specific draft type.

DRAFT	TEAM	FREQ	AVG POT	POT13+
NEW	728	0.975	11.000	0.026
OLD	179	0.926	10.865	0.110
OLD	465	0.909	11.306	0.181
NEW	573	0.950	11.605	0.184
OLD	42	0.915	11.342	0.186

Table 6. Top 5 teams by percentage of draftees with at least 13 draft/current potential by draft type. Draft frequency is the percentage of times that team actually drafted a player under a specific draft type.

DRAFT	TEAM	FREQ	AVG POT	POT13+
NEW	562	1.000	12.700	0.675
NEW	279	1.000	12.600	0.625
NEW	239	1.000	12.400	0.600
NEW	339	0.925	12.432	0.595
NEW	100	1.000	12.675	0.575

considering other contextual factors such as scouting report, player attributes, and amateur/overseas player performance statistics.

By focusing solely on draft potential in making the decision on who to draft, this would therefore artificially inflate their percentage of draftees with at least 13 potential as they would automatically pick these players as long as they were present in their draft pool for that round of the draft.

4. Takeaways

There are a few primary conclusions managers should have from the findings in this paper.

First, the introduction of the new draft system has resulted in a patent inflation of the quality of draft picks adjudged on the basis of player potential. This is primarily due to information asymmetry relative to the old draft system as much more information including the player's potential is now visible prior to making a draft selection. Given this potential can now change, however, it is not enough to only look at player potential in making the decision and other key contextual factors such as player attribute, scouting report, and amateur/overseas performance must also be considered in evaluating player effectiveness. This prospect potential inflation was most evident immediately after the introduction of the new draft system in the 2041 season, and it has now gradually lessened so prospects today have a distribution in potential relatively similar to that under the old draft system.

In addition, managers should pick from the high school draft pool in the first round to maximize their likelihood of receiving a player with at least 13 potential. From the second round onward, the college pool would be the optimal draft pool by the metric of drafting a prospect with at least 13 potential. However, a blanket recommended draft pool is unrealistic due to the changing needs of managers. For those who want draft picks ready to make an immediate impact or receive only minimal training in the minors, the Asian Posting System is recommended since it has the highest draft SI out of all the draft pool for every round. For those who do not have as much depth in the low minors, the high school and Latin Academy draft pools remain good choices to get prospects who are younger and more malleable in terms of training at a specific playing position for example.

Furthermore, managers often seem to be reaching for pitchers in the early rounds of the draft likely due to need. On the sole basis of optimizing the likelihood of selecting a prospect with at least 13 potentials, managers are recommended to focus on the middle infielder pools for the first round, and either the middle infielder or outfielder pools for the second and third. Of course, this would be subject to varying draft needs between teams and overall the difference between positions in receiving a player with at least 13 potential is secondary to the draft source pool, which contains more noticeable differences by this metric between various draft source pools.

Overall, there is significant variance in the quality of draftees by different teams, but for the most part there has been a noticeable uptick in their quality after the immediate implementation of the new draft system, with middle infielders from the college draft pool most likely to have at least 13 potential for all rounds except the first one, which is middle infielders from the high school draft pool.

5. Acknowledgements

This paper represents an update with the new draft system on my previous work. (Luo, 2018) The original inspiration for that report came from another manager in BrokenBat, lostraven, who manually tracked the first round of the 2035 draft. Of course, none of this would be possible without the work of Steve Muller, the creator of BrokenBat.

References

Luo, H. Brokenbat draft report, 2018.

Muller, S. End of the old draft, 2019.

Muller, S. About broken bat, 2020.