What Constitutes a Good Block? An In-Depth Look at the In-Game Effect of a Block in the NBA

Analysis Completed by Kenny Huang and Brian Lin
Report Written by Kenny Huang

May 3, 2022

1 Problem Statement

The field of sports analytics is rapidly growing, and with it comes a challenge to traditional statistics. In basketball, there are seven basic such individual metrics: points, rebounds, assists, steals, blocks, player fouls, and turnovers. Every game has a box score, a detailed table of how many of each metric each player recorded during that game. Until the introduction of advanced metrics in the mid-1990s, most of player evaluation was done by eye, but what little statistical valuation was done was based entirely on these seven numbers. In recent years, analysts have created proprietary, complicated, and frankly confusing metrics like RAPTOR from combinations of the seven as well as a trove of additional data collected for this purpose. However, in this paper, we seek to take a different approach and instead examine the game more closely through an existing statistic: blocks.

Figure 1: A defensive player (6) blocks a shot by an offensive player (22)

By definition, a block is a legal deflection of a shot from an offensive player by a defensive player. As one of the two strictly defensive metrics (along with steals), it is commonly used as a unit of measurement for defensive ability. However, we are arguing that the current definition of blocks is not a good unit of measurement because not all blocks are created equal; as the metric stands today, it doesn't take into account what happens the rest of the current possession— before the other team gets the ball— and the effect on future possessions. In this analysis, we will address both of these limitations and provide an analytical but more approachable look at NBA defense by building on the established statistic of blocks.

2 Related work

This project was directly inspired by a previous article titled "Bad Blocks", researched and written by myself. While the analysis is comprehensive, I felt that there were a few major shortcomings as well as interesting areas of future research. When Brian and I were given the opportunity to complete a sports analytics project in class, we jumped on the opportunity to build on "Bad Blocks" and create a more interesting and clear analysis.

3 Dataset

For the purposes of our possession-based analysis, we needed data at the possession level. Unfortunately, there is no publicly available dataset that fine-grained, so we used the Python BeautifulSoup package to scrape the dataset from a website called Basketball Reference. As can be seen in Figure 2, Basketball Reference (BR) has detailed play-by-play records for every game dating back to 1996.

Play-By-Play
Jump to: 1st | 2nd | 3rd | 4th
scoring play | tie | lead change

1st Q

Time	Boston		Score		Milwaukee
12:00.0	Jump ball: D. Theis vs. B. Lopez (G. Antetokounmpo gains possession)				
11:41.0			0-2	+2	G. Antetokounmpo makes 2-pt hook shot from 2 ft (assist by J. Holiday)
11:22.0	J. Brown makes 3-pt jump shot from 26 ft (assist by D. Theis)	+3	3-2		
11:09.0			3-4	+2	W. Matthews makes 2-pt layup from 2 ft (assist by K. Middleton)
10:49.0	D. Theis misses 2-pt layup from 4 ft		3-4		
10:45.0			3-4		Defensive rebound by B. Lopez
10:38.0			3-4		Personal foul by J. Brown (drawn by B. Lopez)
10:26.0			3-4		G. Antetokounmpo misses 2-pt jump shot from 17 ft (block by G. Williams)
10:22.0	Defensive rebound by G. Williams		3-4		
10:21.0	G. Williams makes 2-pt layup from 1 ft	+2	5-4		
10:12.0			5-6	+2	K. Middleton makes 2-pt layup from 3 ft
9:58.0	Shooting foul by W. Matthews (drawn by J. Brown)		5-6		
9:58.0	J. Brown misses free throw 1 of 2		5-6		
9:58.0	Offensive rebound by Team		5-6		
9:58.0	J. Brown makes free throw 2 of 2	+1	6-6		
9:48.0			6-6		G. Antetokounmpo misses 3-pt jump shot from 26 ft
9:45.0	Defensive rebound by D. White		6-6		
9:33.0	M. Smart misses 2-pt jump shot from 13 ft		6-6		
9:31.0			6-6		Defensive rebound by G. Antetokounmpo
9:21.0			6-6		B. Lopez misses 3-pt jump shot from 26 ft
9:19.0	Defensive rebound by G. Williams		6-6		
9:14.0	J. Brown misses 3-pt jump shot from 26 ft		6-6		
9:12.0			6-6		Defensive rebound by J. Holiday
9:02.0			6-6		J. Holiday misses 3-pt jump shot from 26 ft
8:59.0	Defensive rebound by J. Brown		6-6		
8:55.0	M. Smart makes 3-pt jump shot from 28 ft (assist by J. Brown)	+3	9-6		
8:41.0			9-9	+3	B. Lopez makes 3-pt jump shot from 27 ft (assist by J. Holiday)
8:30.0	J. Brown misses 3-pt jump shot from 27 ft		9-9		
8:27.0			9-9		Defensive rebound by B. Lopez
8:19.0			9-9		B. Lopez misses 2-pt layup from 7 ft
8:18.0			9-9		Offensive rebound by B. Lopez
8:17.0			9-11	+2	B. Lopez makes 2-pt layup from 2 ft
8:07.0	D. Theis makes 3-pt jump shot from 27 ft (assist by J. Brown)	+3	12-11		
7:54.0			12-11		K. Middleton misses 2-pt jump shot from 15 ft
7:51.0	Defensive rebound by G. Williams		12-11		
7:50.0	Shooting foul by J. Holiday (drawn by J. Brown)		12-11		

Figure 2: Sample play-by-play data from Basketball Reference

Using the descriptions, scores, and timestamps, we were able to generate a dataset of the 2021-22 NBA season with our desired features, including the following:

- GameID - Unique Game ID generated by BR of the game the possession belongs to
- away_home - Indicator of whether team in possession was home team
- poss_time - Possession time in seconds
- points_scored - Number of points scored in possession
- num_orebs_team - Number of team offensive rebounds in possession

- num_orebs_ind - Number of offensive rebounds by individual players in possession
- num_1shots - Number of free throws in possession
- num_2shots - Number of two-point shots in possession
- num_3shots - Number of three-point shots in possession
- distance - Distance from basket of first shot in feet
- turnover - Indicator of whether possession ended with turnover
- num_blocks - Number of blocks that occurred in possession
- block_time - Time from block to end of possession in seconds; NA if no block occurred in possession
- block_distance - Distance from basket of blocked shot in feet; refers to first block if multiple blocks occurred in possession; NA if no block occurred in possession
- blocked - Unique Player ID generated by BR of player whose shot was blocked in this possession; refers to first block if multiple blocks occurred in possession; NA if no block occurred in possession
- blocker - Unique Player ID generated by BR of blocking player; refers to first block if multiple blocks occurred in possession; NA if no block occurred in possession
- block_def - Indicator of whether defense secured the ball after the block; refers to first block if multiple blocks occurred in possession; NA if no block occurred in possession

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	GameID	away_home	team	oppt	time	poss_time	block_time	points_scor	num_blocks	num_orebs_	num_orebs_	num_1shots	num_2shots	num_3shots	distance	turnover	bloc
2	202204100N	1	NOP	GSW	48	26		2	0	0	0	0	1	0	14	0	
3	202204100N	0	GSW	NOP	47.5666667	16		3	0	0	0	0	0	1	27	0	
4	202204100N	1	NOP	GSW	47.3	16		0	0	0	0	0	0	1	23	0	
5	202204100N	0	GSW	NOP	47.0333333	18		0	0	0	0	0	0	1	25	0	
6	202204100N	1	NOP	GSW	46.7333333	16		0	0	0	1	0	0	2	26	0	
7	202204100N	0	GSW	NOP	46.4666667	13		0	0	0	0	0	0	1	27	0	
8	202204100N	1	NOP	GSW	46.25	4		2	0	0	0	1	1	0	2	0	
9	202204100N	0	GSW	NOP	46.1833333	10		1	0	0	0	0	0	0		1	
10	202204100N	1	NOP	GSW	46.0166667	5		2	0	0	0	0	1	0	0	0	
11	202204100N	0	GSW	NOP	45.9333333	20		0	0	0	0	0	0	1	23	0	
12	202204100N	1	NOP	GSW	45.6	1		3	0	0	0	1	1	0	2	0	
13	202204100N	0	GSW	NOP	45.5833333	13		2	0	0	0	0	1	0	15	0	
14	202204100N	1	NOP	GSW	45.3666667	15		3	0	0	0	0	0	1	23	0	
15	202204100N	0	GSW	NOP	45.1166667	19		2	0	0	0	0	1	0	13	0	
16	202204100N	1	NOP	GSW	44.8	15		0	0	0	1	0	0	1	25	1	
17	202204100N	0	GSW	NOP	44.55	13		2	0	0	0	0	1	0	17	0	
18	202204100N	1	NOP	GSW	44.3333333	21		0	0	0	0	0	0	1	22	0	
19	202204100N	0	GSW	NOP	43.9833333	10		3	0	0	0	0	0	1	25	0	
20	202204100N	1	NOP	GSW	43.8166667	24		2	0	0	2	0	3	0	2	0	
21	202204100N	0	GSW	NOP	43.4166667	19		0	0	0	0	0	0	1	25	0	
22	202204100N	1	NOP	GSW	43.1	6		2	0	0	0	2	0	0		0	
23	202204100N	0	GSW	NOP	43	16		0	0	0	0	0	0	1	24	0	
24	202204100N	1	NOP	GSW	42.7333333	11		0	0	0	0	0	0	1	26	0	
25	202204100N	0	GSW	NOP	42.55	12		3	0	0	0	0	0	1	25	0	

Figure 3: Sample possession data scraped using BeautifulSoup

3.1 Limitations of Dataset

For simplicity and consistency, we did not include overtime in our dataset. We admit that it would be interesting to look at overtime possessions specifically and compare the results to regulation possessions, but because the scope of this analysis is only the 2021-22 season, the sample size would be too small for us to gain any robust insights. In addition, we would point out that our dataset draws entirely from the BR database. As amazing at BR is, there are some accuracy concerns for such a large open-source project, although the inaccuracies that we found were scarce and uncorrelated with each other.

4 Analysis: Redefining Block Evaluation for Individuals

Our basic premise for this analysis is that the goal of defense is to prevent points. As such, "good" blockers should not be defined as players who accumulate large numbers of blocks but rather should be those who minimize the number of points scored by the opposing team. Currently, defensive rating, defined as the average number of points scored by the opposing team per 100 possessions, is already commonly used, but today we will be modifying this by conditioning on possessions containing a block. That is, on possessions that Player A blocked a shot, how many points did the opposing team score anyway? If on average the opposing team still scores often, then Player A would not be considered a "good" blocker, even if they block a large number of shots.

But what's the point? What does conditioning on blocked possessions actually do for us? The answer is largely for perception correction. Through the "Bad Blockers" article, we realized that many players traditionally thought of as great defenders and blockers turned out to have given up ridiculous numbers of points anyway. In contrast, underrated defenders like Greg Ostertag and Chris Webber excelled in this area and deserve more praise for their defensive efforts. Renowned greats like Tim Duncan reinforced their images by passing both the eye test and numbers test with flying colors. By generating a more detailed evaluation metric in the specific domain of blocks, we provide a tool that can be used to find undervalued gems and avoid empty stats. While this analysis as it stands doesn't have many in-game applications, we hope that our work can contribute to the body of work of defensive evaluation, an area that is steadily gaining visibility but is still far too murky.

4.1 Some New Metrics

For this analysis, we defined a player's points per block (PPB) as the expected points scored in possessions where they recorded a block. This can be broken down into two new metrics: offensive rebounding rate (ORR)— the proportion of their blocks that were rebounded by the offense— and points per offensive rebound (PPO)— the expected points scored in possessions after an offensive rebound. Mathematically, because points can only be scored if the block is followed by an offensive rebound, PPB is always equal to the product of ORR and PPO. In addition, we also computed the points blocked (PBK) by each player. This is defined as the expected points scored by the specific shots that they blocked.

4.2 Blocking Efficiency: ORR and PPO

As a whole, the league had an average ORR of about 41% and an average PPO of 1.07. That is, 41% of the time, the ball ended up back with offense, who scored an average of 1.07 points in those possessions. For reference, offenses scored an average of 1.12 points overall. Intuitively, these make sense: a majority of blocks are by primary rim protectors, who are taller and out of position after the block to get the defensive rebound, which explains why the offensive rebounding number is higher than the normal average of 27%. After a block, there is also less time on the shot clock for the offense to shoot if the ball leaves the paint area, resulting in less efficient offense for the remainder of the possession that offsets any benefit from the defender being out of position. In fact, if the ball goes out of bounds and everything resets, the offense with the additional time pressure only scores 0.96 points, compared to 1.19 if the ball is controlled by an offensive player.

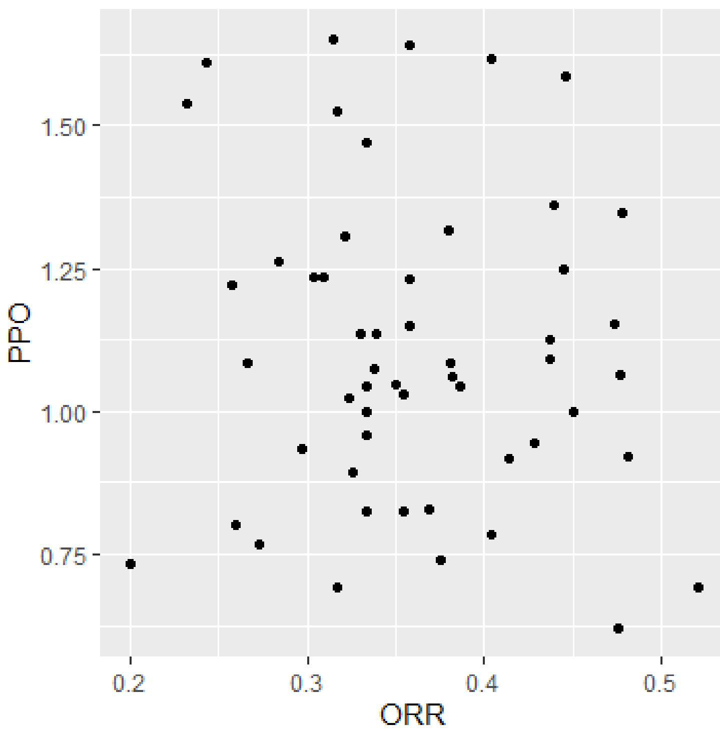

Figure 4: Plot of ORR vs PPO for players with minimum of 50 blocks in 2021-22 season

Interestingly, there is a negligible relationship between the two factors, with a correlation coefficient of 0.053, meaning that we can treat them as orthogonal when ranking them later. When looking at the 53 players who had at least 50 blocks this past season, there's a strong negative correlation of height with both ORR (-0.36) and PPO (-0.59), with slightly weaker correlations of weight with ORR (-0.34) and PPO (-0.27). In the context of blocking efficiency, Isaiah Stewart (DET), Kristaps Porzingis (WAS), JaVale McGee (PHX) stand out.

5 Blocking Volume: PBK

Thus far, the metrics that we discussed were purely from an efficiency standpoint. However, even accounting for outliers by only including players with at least 50 blocks, we could not definitively say the most efficient blockers are thus the "best" blockers. In our opinion, volume does matter as well. However, to more accurately gauge the scale of a player's impact, we computed a new metric called Points Blocked that is the expected points scored by their blocked shots had they not been blocked. To do this, we made the simplifying but reasonable assumption that the primary determinant was the distance from the basket and computed the average points scored by shots from each distance.

For each blocked shot, we then found the average points scored with similar shots, and a player's points blocked is the sum of all of those values. This gives a better indication of the volume impact than the raw number of blocks. Most of the rankings stay the same, but Robert Williams (BOS)

Distance			
0	1.832	18	1.006
1	1.739	19	1.000
2	1.536	20	0.952
3	1.354	21	0.922
4	1.220	22	1.154
5	1.138	23	1.353
6	1.110	24	1.347
7	1.137	25	1.270
8	1.092	26	1.241
9	1.128	27	1.235
10	1.099	28	1.231
11	1.102	29	1.148
12	1.095	30	1.167
13	1.103	31	1.092
14	1.052	32	1.029
15	1.031	33	0.955
16	1.030	34	0.759
17	1.015	35	0.825

Figure 5: Expected points scored in possession by distance of shot

and Myles Turner (IND) were leap-frogged by Jakob Poeltl (SAS) and Evan Mobley (CLE) who had less blocks but a higher number of points blocked. Overall, Jaren Jackson Jr. widely leads the field, followed by Rudy Gobert and Mitchell Robinson.

Player	PBK	Count
Jaren Jackson Jr.	228.54	166
Rudy Gobert	177.02	136
Mitchell Robinson	175.47	131
Jakob Poeltl	157.05	112
Evan Mobley	156.43	110
Myles Turner	155.95	117
Robert Williams	155.08	126
Mo Bamba	139.83	112
Joel Embiid	130.9	96
Hassan Whiteside	130.89	101

Figure 6: Highest volume blockers

As a final note, we also checked the relationship between the efficiency metrics and PBK. Among the players with at least 50 blocks, the correlation coefficients were all 0.20 or less, indicating a weak at best relationship and confirming the validity of creating these two separate categories.

6 Overall Rankings

Taking our efficiency and volume rankings together, we can obtain our overall rankings. Although there are many ways to approach this, we decided on a geometric mean approach. To compute this, we first found all 53 players who had at least 50 blocks during the 2021-22 season and ranked them in each of ORR, PPO, and PBK. Next, we took the geometric mean of the ORR and PPO ranks to obtain the efficiency rank score, and then took the geometric mean of that and the PBK rank, representing the volume criterion, to obtain the final rank score. We thus deemed the players with the lowest final rank scores to be the best overall blockers during this season.

Rank	Player	ORR_Rank	PPO_Rank	PBK_Rank	Overall
1	Jaren Jackson Jr.	41	16	1	5.06
2	Evan Mobley	7	6	5	5.69
3	Rudy Gobert	16	20	2	5.98
4	Isaiah Stewart	1	4	22	6.63
5	Mitchell Robinson	36	25	3	9.49
6	Jakob Poeltl	15	42	4	10.02
7	Kristaps Porzingis	14	2	19	10.03
8	Hassan Whiteside	9	15	10	10.78
9	JaVale McGee	5	8	21	11.52
10	Myles Turner	26	24	6	12.24

Figure 7: Best overall blockers

In our analysis, Jaren Jackson Jr. is still our blocking champion, barely edging out Evan Mobley and Rudy Gobert who round out our top 3.

7 Analysis: Quantifying Effect on Future Possessions

One of the most commonly asked questions about the "Bad Blockers" article was about momentum. With game-changing blocks like LeBron's in the 2016 NBA Finals crystal clear in our memory, we hypothesized that a block positively impacts the game in favor of the team that recorded the block. In other words, blocks shift the momentum in the favor of the defending team in a way that can be numerically measured in the ensuing possessions.

To test this hypothesis, we computed each team's average points per possession for each game. In addition, for each block, we computed the average points scored per possession in the next few possessions for both teams. Finally, we computed the following values for $x = 1, 2, 3, 4, 5$:

$$\text{blocking team: average PPP over next } x \text{ possessions} - \text{average PPP over entire game}$$
$$\text{blocked team: average PPP over entire game} - \text{average PPP over next } x \text{ possessions}$$

We expect both metrics to be negative: the momentum shifts to the blocking team at the expense of the blocked team. However, once we computed them, we found the exact opposite to be true. Not only were the averages for both metrics positive for all x, it was significantly positive at the 95% level. Contrary to what we thought, teams scored an average of 1.6% higher after being blocked. When we tried to explain why, we came up with the possible hypotheses of shot selection, three-point shooting, and free throw shooting.

7.1 Shot Selection

Our first hypothesis was that after being blocked, players might be more hesitant to attack the rim again and will instead shoot more midrange or three-point shots. After examining the shot distributions of the blocked shots and the shots in the possession immediately after, we concluded that this was almost surely not the cause. Aside from there being little difference visually (Figure 8), we also tested the hypothesis using t-tests and found no significant differences for any distance rounded to the nearest foot.

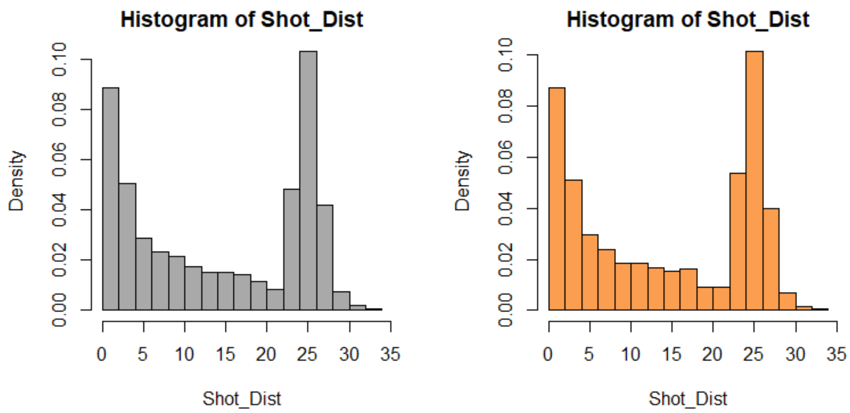

Figure 8: Two shot distributions

7.2 Three-Point Shooting

Another hypothesis more specifically targeted three-point shooting. However, there were similar results here to the first hypothesis. In the five possessions before blocks, 1.673 of them involved a three-point shot, compared to 1.650 of the five possessions immediately after, a negligible difference. We also tried to visualize this using a two-step matrix, where the row indicates the frequency in the previous possessions and the column indicates the frequency in the following possessions. Generally, if the hypothesis were true, this would be reflected in the matrix as having larger numbers in the top right half of the matrix in Figure 9, which is not the case.

		Following possessions with 3-pointer					
		0	1	2	3	4	5
Previous possessions with 3-pointer	0	231	551	504	253	60	7
	1	517	1173	1165	527	135	19
	2	537	1145	1122	605	176	18
	3	254	616	604	308	99	9
	4	64	161	176	100	20	2
	5	5	14	21	8	2	0

Figure 9: Two-step matrix of three-point outcomes

7.3 Free Throw Shooting

Our last hypothesis to test was that after blocks, players draw more fouls due to a combination of more aggressive play, crowd pressure on referees, and possibly other factors. To our surprise, when we computed how many of the five possessions before or after the block involved free throws, the numbers rose from 0.5534 before to 0.5879 after. Statistically, our t-test yielded a z-score of 5.2, correlating with a p-value of 4.8×10^{-7}. A two-step matrix analogous to the one made for three-pointers verifies this finding; the numbers above the diagonal are indeed higher.

Previous possessions with free throws \ Following possessions with free throws	0	1	2	3	4
0	3522	2130	546	95	6
1	2049	1246	402	58	10
2	490	368	117	28	3
3	67	42	12	5	1
4	4	3	4	0	0

Figure 10: Two-step matrix of free-throw outcomes

This suggests that there is some truth to this theory. It would help to look deeper into this to see what exactly is causing this difference, but this is outside the scope of this analysis.

8 Future Work

As much as we believe in our analysis, there are some limitations or avenues of future research. First, we would like to analyze further for late game and overtime possessions specifically, as these will be the most interesting. Also, the points blocked metric could be improved by including more features in the model than just the distance. If more data is available about closest defender, etc., this could be improved even further. Finally, there's so many different ways that the final rankings could be produced, so there could be a better way that makes more sense.

9 Acknowledgements

We would like to thank Professor Alain Kornhauser for allowing us this opportunity and providing us with feedback and suggestions. In addition, we would like to thank Basketball Reference for the amazing work they do for sports nerds like ourselves and giving us the possibility to complete this project in the first place.

References

[1] Gómez, Emilia et al. "Visual Music Transcription of Clarinet Video Recordings Trained with Audio-Based Labelled Data". 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). N.p., 2017. 463–470. Web. https://ieeexplore.ieee.org/document/8265272

[2] Zhao, Hang et al. "The Sound of Pixels". CoRR abs/1804.03160 (2018): n. pag. Web. https://arxiv.org/abs/1804.03160

[3] https://docs.opencv.org/3.4/d9/db0/tutorial_hough_lines.html

[4] https://stackoverflow.com/questions/45075638/graph-k-nn-decision-boundaries-in-matplotlib