Paper B2B-03

Basketball Analytics: Optimizing the Official Basketball Box-Score (Play-by-Play)

William H. Cade, CADE Analytics, LLC

ABSTRACT

What if basketball analytics could formulate an "**end-all**" value that could justly evaluate team and/or player performance? Those immersed in the world of basketball analytics are challenged with this mission: to translate a game of interdependent factors into simple measures of player and team performance. The **Official Basketball Box-Score (Play-by-Play)**, for example, significantly serves as an invaluable guide in understanding the analysis of the game on a fundamental level. With an unbiased precision, the **Official Basketball Box-Score (Play-by-Play)** displays descriptive measurements used to inform how well or how poorly a player and/or team have performed. With the usage of advanced statistical software tools, I have extended the **Official Box-Score** and established a never-ending framework that can measure the offensive and defensive prowess for a basketball team, by lineup (overall and per game). Relative to these measurements of frequency, efficiency, and precision, I will discuss its methodology and display examples of output for the evaluation of **player** and/or **team lineup** performance. Though basketball analytics comes with its limitations and imperfections, the pursuit of the advancement of knowledge of the game further incites ongoing analyses and a penchant for **better statistics!**

INTRODUCTION

In the game of basketball, analytics are challenged with the mission to translate a game of interdependent factors into simple measures of player and/or team performance. The "four factors" established by Dean Oliver's *Basketball on Paper*, have long played a role in the understanding and evaluation of team success:

1. Shooting
2. Turnovers
3. Rebounding
4. Free Throws

Simply put: If you're better at your opponent at making field goals, creating turnovers, grabbing rebounds, and getting to the foul line, then you're going to win many more games than you will lose. By utilizing the collection of information provided by the Official Basketball Box Score (Play-by-Play) "**PEP-BALL**" (**P**erformance **E**valuation of **P**layer(s)), developed by CADE Analytics, LLC, is a method powered by BASE SAS® used to evaluate player and/or team performance. By statistically identifying the qualities shared by players and opponents, **PEP-BALL** looks at various combination of players on the court and shows which combinations work the best-the most effective two-player, three-player and even five-player combinations for each game, and more importantly, each game situation.

Official Basketball Box Score (Play-by-Play)

The Official Basketball Box-Score (OBBS) serves as an invaluable guide in understanding the analysis of the game on a fundamental level. With an unbiased precision, the composition of the OBBS is two-fold. First, it presents descriptive measurements used to inform how well or how poorly a player and/or team have performed (Figure 1).

Official Basketball Box Score -- Game Totals -- Final Statistics
Michigan vs Louisville
04/08/13 9:23 p.m. at Georgia Dome, Atlanta, GA

Michigan 76 • 31-8

##	Player		Total FG-FGA	3-Ptr FG-FGA	FT-FTA	Rebounds Off	Def	Tot	PF	TP	A	TO	Blk	Stl	Min
01	Robinson III, Glenn	f	3-4	0-1	6-8	1	1	2	1	12	2	0	0	0	38
04	McGary, Mitch	f	3-6	0-0	0-0	2	4	6	4	6	1	1	1	1	29
03	Burke, Trey	g	7-11	3-5	7-9	1	3	4	4	24	3	4	1	0	26
10	Hardaway Jr, Tim	g	5-13	0-4	2-4	0	5	5	0	12	4	2	0	0	35
11	Stauskas, Nik	g	1-2	1-2	0-0	0	2	2	3	3	2	1	0	1	19
02	Albrecht, Spike		6-9	4-5	1-2	0	1	1	1	17	0	3	0	0	28
15	Horford, Jon		0-0	0-0	0-0	2	0	2	1	0	0	0	0	0	5
23	LeVert, Caris		0-1	0-1	0-0	1	2	3	1	0	0	1	0	1	12
52	Morgan, Jordan		0-2	0-0	2-2	1	0	1	0	2	0	0	0	0	8
	Team					0	1	1							
	Totals		25-48	8-18	18-25	8	19	27	15	76	12	12	2	3	200

FG %	1st Half:	14-28	50.0%	2nd half:	11-20	55.0%	Game:	25-48	52.1%
3FG %	1st Half:	6-11	54.5%	2nd half:	2-7	28.6%	Game:	8-18	44.4%
FT %	1st Half:	4-6	66.7%	2nd half:	14-19	73.7%	Game:	18-25	72.0%

Deadball Rebounds: 5

Louisville 82 • 35-5

##	Player		Total FG-FGA	3-Ptr FG-FGA	FT-FTA	Rebounds Off	Def	Tot	PF	TP	A	TO	Blk	Stl	Min
20	BLACKSHEAR, Wayne	f	3-5	2-3	0-0	0	1	1	3	8	1	0	0	1	19
21	BEHANAN, Chane	f	6-12	0-0	3-4	7	5	12	2	15	1	2	0	1	28
10	DIENG, Gorgui	c	4-6	0-0	0-0	5	3	8	4	8	6	2	3	1	37
02	SMITH, Russ	g	3-16	1-6	2-3	1	0	1	4	9	2	3	0	0	32
03	SIVA, Peyton	g	6-15	0-2	6-6	0	6	6	3	18	5	2	0	4	36
11	HANCOCK, Luke		5-6	5-5	7-10	0	1	1	4	22	3	0	0	2	30
15	HENDERSON, Tim		0-0	0-0	0-0	0	0	0	1	0	0	0	0	0	3
24	HARRELL, Montrezl		1-1	0-0	0-0	0	0	0	1	2	0	0	0	0	12
44	VAN TREESE, Stephan		0-0	0-0	0-0	2	0	2	0	0	0	0	0	0	3
	Team					0	1	1							
	Totals		28-61	8-16	18-23	15	17	32	22	82	18	9	3	9	200

FG %	1st Half:	12-26	46.2%	2nd half:	16-35	45.7%	Game:	28-61	45.9%
3FG %	1st Half:	5-8	62.5%	2nd half:	3-8	37.5%	Game:	8-16	50.0%
FT %	1st Half:	8-9	88.9%	2nd half:	10-14	71.4%	Game:	18-23	78.3%

Deadball Rebounds: 4

Figure 1. NCAA 2013 Championship Game OBBS- Descriptive Measurements

Descriptive measurement statistics included in an OBBS are field goals made and field goals attempted (“FGM” and “FGA”), three-point field goals made and three-point field goals attempted (“3PM” and “3PA”), free throws made and free throws attempted (“FTM” and “FTA”), offensive rebounds (“OR” or “OREB”), defensive rebounds (“DR” or “DREB”), total rebounds (“TREB”), assists (“A” or “AST”), steals (“S” or “STL”), blocked shots (“B” or “BS”), personal fouls (“F” or “PF”) , Turnovers (“TOV” or “TO”), minutes (“M” or “MIN”) and points (“P” or “PTS”).

Secondly, the play-by-play segment (Figure 2**)** of the OBBS provides a sequential detailed description of events, with information relative to time and active players. More importantly, it provides a documented account of game play, respective to specific players (**team lineups**) in the game.

Michigan vs Louisville
04/08/13 9:23 p.m. at Georgia Dome, Atlanta, GA
1st PERIOD Play-by-Play (Page 2)

HOME TEAM: Louisville	Time	Score	Margin	VISITORS: Michigan
	11:41			TIMEOUT MEDIA
	11:41			SUB IN : McGary, Mitch
	11:41			SUB OUT: Horford, Jon
REBOUND (DEF) by HANCOCK, Luke	11:20			MISSED JUMPER by McGary, Mitch
MISSED FT SHOT by HANCOCK, Luke	11:09			FOUL by Burke, Trey (P2T3)
REBOUND (OFF) by (DEADBALL)	11:09			
GOOD! FT SHOT by HANCOCK, Luke	11:09	16-20	V 4	
GOOD! FT SHOT by HANCOCK, Luke	11:09	17-20	V 3	
SUB IN : VAN TREESE, Stephan	11:09			SUB IN : LeVert, Caris
SUB OUT: DIENG, Gorgui	11:09			SUB OUT: Burke, Trey
	10:53			TURNOVR by Hardaway Jr, Tim
STEAL by HANCOCK, Luke	10:52			
TURNOVR by SMITH, Russ	10:33			
	10:11			TURNOVR by Hardaway Jr, Tim
STEAL by SIVA, Peyton	10:09			
MISSED LAYUP by SMITH, Russ	10:06			
REBOUND (OFF) by VAN TREESE, Stephan	10:06			
SUB IN : HENDERSON, Tim	09:46			SUB IN : Stauskas, Nik
SUB IN : DIENG, Gorgui	09:46			SUB OUT: Hardaway Jr, Tim
SUB OUT: VAN TREESE, Stephan	09:46			
SUB OUT: SIVA, Peyton	09:46			

Figure 2. NCAA 2013 Championship Game OBBS- Play-by Play

A fundamental building block for the measure of team performance is time, as function of minutes (“playing time”). Within the analytics of basketball, playing time is sequenced and coded as a “possession.” A possession in basketball is defined as the period of time between when one team gains control of the basketball and when the opposing team gains control of the basketball. Contrast with play. How much information observed and collected on a particular possession is fundamental in identifying the basketball “DNA” associated with/within a team (player and/or lineup). This process of collecting valued information is defined as **game-charting** and is displayed in the following truncated example (Figure 3):

A	B	C	D	E	F	G	K	L	M	P	Q	R	S
date	opponent	half	time_rem	d_line_up	poss	team_off	o_shot_out	o_score	team_def	d_shot_out	d_score	line-up	misc
######	########	1	19.36	0,1,4,35,45	1					MFG (42), ORB (5), MFG (5), DRB (45)		10,22,42,5,13	
					2		M3FG (35), D3RB (42)			MFG (42), DRB (4)			
					3		FG (35) -A (0)	2-0		FG (22)	2-2		
					4		MFG (1), DRB (13)			MFG (13), DRB (4)			
					5		FG (35) -A (0)	4-2		MFG (13), ORB (22), FG (42) -A (10)	4-4		
					6		TOV-STEAL (35)			M3FG (10), D3RB (35)			
					7		M3FG (4), D3RB (13)			TOV-STEAL (4)			
			15.51		8		TOV-BOB (4)						
					8				Y (45)	FT (22)	4-6	22,5,13,54,0	
				0,1,10,15,23	9		TOV-STEAL (1)			FG (23)	4-8	22,12,13,54,0	
					10		BFG (1), TDRB						
				0,4,10,15,23	10					MFG (5), ORB (5), MFG (0), DRB (0)		22,5,13,54,0	

Figure 3. Excel Spreadsheet Game-Charting

Validated by the aforementioned Play-by-Play (Figure 2) component of the OBBS, game-charting establishes the infrastructure for the evaluation of player and/or team (lineup) performance

SAS METHODOLOGY FOR BASKETBALL ANALYTICS

PROC IMPORT

The IMPORT procedure reads the external basketball game data (Figure 2) and writes the data to a SAS data set. A simple macro is invoked within the analytic programming code such that seasonal basketball data is accumulated by subsequent macro parameters. For example, the following macro named %BASKETBALL has parameters for *opponent* (team name), *range* (EXCEL spreadsheet data cell range), and *name* (player name):

```
%macro basketball (opponent, range, name);

proc import out= work._&opp datafile= "C:\game_charting.xls"
                    dbms=excel2000 replace;
       range=&range;
       sheet=&opp;
       getnames=yes;
run;


%mend basketball;
%importer (opp=USF, range="A1:S82", oppname=SOUTH_FLORIDA.);
```

The IMPORT procedure, along with the macro provides **PEP-BALL** with the platform to examine the historical performance of a team and its opponent (per game, seasonally, etc.)

PROC FORMAT

The two main areas of display in Figure 1 are the basketball players (**team lineups**) and their respective basketball statistics. The %INCLUDE describes the location of the information that enables **PEP-BALL** to define a player's jersey number with their respective names:

```
%include "C:\ bball_format.sas";
```

As a result, the PROC FORMAT creates a user-defined format pertinent to a player's respective team (bball_format.sas):

```
proc format;
 value $name
       10="Thomas"
       22="Everett"
       42="William"
       5="Paige"
       13="Jacobs"
       0="Ward"
       1="Jackson"
       12="A. Jordan"
       25="Wingate"
       54="Dez"
       ;
run;
```

DATA Step

The DATA step executed in the "box-score" programming code permits true programming functionality, in relation to basketball analytics:

```
data
       _&opp;

       set _&opp;
run;
```

By adding the SET statement, which specifically points to the imported SAS data set to be read (i.e. game_charting.xls), the DATA step collects and augments information in the following variety of ways: *retain information*, *create accumulating totals*, and *summarize data sets*

RETAIN Information:

The RETAIN statement appears within the DATA step to initialize and retain fixed variables of interest. For example, **team lineups** are uniformly a significant variable of interest in the evaluation of team performance.

```
*RETAIN LINEUP;
if lineup ^= ' ' then save=lineup;
else lineup=save;
retain save;
drop save;
```

In doing so, the imported SAS data set is sorted appropriate to the variable, **lineup.** Retaining team lineup information clearly identifies the qualities shared by players whose play on the court seems simply to flow (or not):

```
proc sort data=_&opp out=_&opp_sorted;
by
       lineup ;
run;
```

Particularly, the application of the SORT procedure sets the foundation of evidence of historical performance, accordant to **team lineups** played.

Create Accumulating Totals

All **descriptive measurement statistics** included in the OBBS are initialized and accumulated in "box-score" programming code. In view of the vast number of repetitive calculations of descriptive statistics, this paper will illustrate only two examples of DO loop processing that generates basketball team data, while eliminating redundant code. For example, the following code generates data that counts the cumulative total of field goals made, by a team:

```
* COUNT OFFENSIVE FIELD GOALS;
cum_fg=0;
i=0;
count_fg=0;
       do until(fg eq " ");
       i+1;
       fg=scan(o_shot_out,i,", ");
       if trim(left(fg)) in ("FG" "3FG" "AND1") then count_fg+1;
       end;
cum_fg+count_fg;
```

Statements within the DO loop execute for a number of iterations until a specific condition (i.e. blank space indicated by " ") stops the loop. The nested SCAN function within the DO loop extracts all field goal related words (FG, 3FG, AND1) defined in the game-charting character variable (o_shot_out). Subsequently, the cumulative totals of field goals made are defined as the numeric variable, **cum_fg**. For all remaining descriptive measures included in the OBBS, the ensuing COUNTER variables parallel the aforementioned DO loop processing code. The following "box-score" code illustrates an additional COUNTER variable used to generate data that counts the cumulative total of defensive rebounds (**cum_drb**) collected, by team:

```
* COUNT DEFENSIVE REBOUNDS;
cum_drb=0;
d=0;
count_drbs=0;
       do until(def_reb eq " ");
       d+1;
       def_reb=scan(d_shot_out,d,", ");
       if trim(left(def_reb)) in ("DRB" "D3RB" "TDRB" "FTDRB" "TD3RB") then
       count_drbs+1;
              end;
cum_drb+count_drbs+(count(d_shot_out,"DEAD3BALLD-RBD"));
```

Interestingly, the relationship between the collections of information (while game-charting) and establishing variables that define all OBBS descriptive measures extend the OBBS in four principles areas. The four principle areas of extension of the OBBS include possession, field goals, rebounding, and free-throw.
***Note:** CADE Analytics, LLC has derived the statistic entitled the "Three-Bound" which is defined as a three-point field goal attempted rebound ("3REB " or "3RB"). Moreover, additional derivative and unique (*) statistics provided in extended OBBS are three-point field goal attempted offensive rebounds ("O3REB" and "O3RB") and three-point field goal attempted defensive rebounds ("D3REB" or "D3RB"), two-point field goals made and two-point field goals attempted ("2FGM" and "2FGA"), offensive possessions ("OP"), defensive possessions ("DP"), free throw offensive rebound ("FTOREB" or "FTORB"), free throw defensive rebound ("FTDREB" or "FTDRB") and potential free throws ("PFT").

Summarize Data Sets
In essence, the collection of information provided in the extended OBBS statistics allows for an efficacious way of showing the best-assembled/best combination of players on the court. Complimentary to the notable "plus/minus" basketball statistic that looks at the point differential when players are both in and out of the game, **PEP-BALL** effectively measures the offensive and defensive aptitude for a basketball team, by lineup (per game, seasonally, etc.).

The example in Figure 4, entitled "Factors of Evaluation" illustrates team lineup performance for the entirety of single basketball game:

Table 1

	Possessions		eFG%		Turnover %		Rebounding%		Assist Rate		Team Foul Rate	
	Off	Def	Off	Def	Off	Def	Off	Def	Off	Def	Off	Def
Thomas, Everett, William, Paige, Jacobs	16	15	50	46.4	25	0	50	33.3	25	50	32.3	19.4
Thomas, Everett, William, Jackson, Jacobs	12	11	27.3	27.8	25	9.1	36.4	42.9	33.3	100	17.4	34.8
Thomas, Everett, Wingate, Paige, Jacobs	8	8	50	62.5	25	50	50	0	33.3	50	37.5	12.5
Everett, William, Jackson, Ward, Jacobs	6	6	50	41.7	0	33.3	50	60	25	100	33.3	16.7
Thomas, Wingate, Ward, Paige, Jacobs	6	6	62.5	62.5	16.7	0	0	0	50	50	0	33.3
Thomas, Wingate, Jackson, Ward, Paige	5	6	50	66.7	20	16.7	66.7	50	100	66.7	0	18.2
Thomas, Everett, Jackson, Ward, Jacobs	3	2	0	0	0	0	0	50	0	0	80	0
Thomas, Everett, Ward, Paige, Jacobs	3	3	0	83.3	0	33.3	100	33.3	0	0	66.7	0
Thomas, William, Wingate, Paige, Jacobs	2	2	125	0	0	50	0	0	100	0	0	50
Everett, Wingate, Jackson, Ward, Paige	2	3	0	62.5	50	0	0	66.7	0	0	0	40
Thomas, William, Jackson, Ward, Jacobs	1	1	0	100	0	0	0	0	0	0	100	100
TOTAL Possessions	64	63										

Figure 4. Factors of Evaluation Example

Identified by what is defined as the "Factors of Evaluation" of team performance, the following provides information on how these 'factors' are measured:

Shooting: The shooting factor is measured using Effective Field Goal Percentage (eFG%). The formula for both offense and defense is (FG + 0.5 * 3P) / FGA. Effective Field Goal Percentage; the formula is (FG+ 0.5 * 3FG) /FGA. This statistic adjusts for the fact that a 3-point field goal is worth one more point than a 2-point field goal.

Turnovers: The turnover factor is measured using Turnover Percentage (TOV%). The formula for both offense and defense is TOV / (Possessions played).

Rebounding: The rebounding factor is measured using Offensive and Defensive Rebound Percentage (ORB% and DRB%, respectively). The formula for offense is ORB / (ORB + Opp DRB), while the formula for defense is DRB / (Opp ORB + DRB).
Assist: The assist factor is measured using Assist Percentage (AST%). The formula is (AST/FGM) for both offense and defense

Team Foul Rate: The foul rate factor is measured using Team and Opponent Foul Percentage. The formula for a team is (Team Fouls/Possession Played), while the formula for the opposing team is (Opponent Team Fouls/Possessions Played).

CONCLUSION

PEP-BALL's impartial approach to quantify team chemistry evaluates a variety of combinations of players on the court and clearly shows which combinations have the biggest effect – best impactful two-player, three-player and even five-player combinations for each game played. **PEP-BALL** is an essential tool for all basketball teams during regular and post-season play. For example, teams invited to the NCAA Tournament evaluate regular- season play and most-recent conference tournament play as a guide for preparation. With the **PEP-BALL** approach, coaches of tournament teams meticulously prepare their players for any situation that may arise in upcoming games. Backed with statistics, coaches can make better decisions and implement effective strategies for both defensive and offensive scenarios. **PEP-BALL** is flexible and can be adjusted to meet any criteria determined by a coach.

Picture this: your team is in the middle of a Sweet 16 game, down by five points with two minutes until the final buzzer. You want to put your best defenders in the game to keep your opponent (who has the ball) from scoring and to increase the chance of a turnover. The **PEP-BALL** method may consider the following data points:

- Which players have the greatest effect on team lineup performance (offensively and/or defensively)?
- Which lineup best defends a two-point field goal attempt?
- Which lineup is best at grabbing the rebound after a missed three-point field goal?

Say you're in the Elite Eight, tied with 30 seconds left to play and you have the ball. The **PEP-BALL** method would consider things like:

- Which lineup fouls the least, yet scores the most efficiently?
- Which lineup has the best assist rate?
- In which particular lineup does your star three-point shooter perform the best?

PEP-BALL incorporates these kinds of situational moments – and many more – from a team's full season using box score analytics. Simply put, the application of **PEP-BALL** provides respective coaches and coaching staffs, with an exact pulse on the effectiveness of their basketball team. By establishing a never-ending framework that can measure the offensive and defensive readiness for a basketball team by lineup (per game, seasonally, etc.), **PEP-BALL** examines the historical performance of a team and its opponent by evaluating the lineups played (offense) and lineups played against (defense).

REFERENCES

Oliver, Dean. 2003. *Basketball on Paper*. pg. 63. Washington, D.C. Potomac Books, Inc.

CONTACT INFORMATION

Name: William Cade
Enterprise: CADE Analytics, LLC
Address: 6715 Miami Lakes Drive C-310
City, State ZIP: Miami Lakes, Florida 33014
Work Phone: (919) 260-8343
E-mail: qstatscade@gmail.com
Web: www.cadeanalytics.com