CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Liberty University Digital Commons

Seeking Excellence: Improving Objectivity in Player Analysis in Professional Basketball

Nathan Cook

A Senior Thesis submitted in partial fulfillment
of the requirements for graduation
in the Honors Program
Liberty University
Fall 2018

Acceptance of Senior Honors Thesis

This Senior Honors Thesis is accepted in partial fulfillment of the requirements for graduation from the Honors Program of Liberty University.

____
David Schweitzer, Ph.D.
Thesis Chair

____
Timothy Van Voorhis, Ph.D.
Committee Member

____
Phillip Blosser, Ph.D.
Committee Member

____
James H. Nutter, D.A.
Honors Director

____
Date

Abstract

This thesis details the creation and testing of an original statistical metric for analyzing individual basketball players in the National Basketball Association (NBA) by both their commonly measured statistics and their so-called "intangibles." By using existing methods as both guides and a caution against potential shortcomings, an inclusive statistic with multiple layers of data can be built to best reflect an individual player's overall value to his team. This metric will be adjusted to account for the differences across multiple eras of NBA play and the levels of talent with which a player played in order to avoid penalizing a player for the unique aspects of his career.

Seeking Excellence: Improving Objectivity in Player Analysis in Professional Basketball

The use of analytics in basketball has grown drastically in recent years as teams have joined an arms race to win games using facts more than opinions. However, the influx of analytics has introduced a new array of difficulties for coaches and strategists in the game. As professional basketball teams have more resources, and thus more access to analytics, than lower level teams, they are also faced with more problems inherent to sifting through the statistics they use. Perhaps the greatest problem is determining the relation between a given statistic for an individual player and the overall value that player has to the team. Solving this problem can be broken down into a few simple steps – understanding the statistic and its peripheral factors, establishing a general measure for overall value, and finding the connection between the initial statistic and that general measure. This requires an in-depth study of commonly used metrics and an application of them toward a more holistic understanding of basketball as quantitative over qualitative.

Overview

Expert Analysis

In professional basketball, every player impacts his team in different ways, making it difficult to measure and compare the values of players in an objective and accurate manner. In order to best eliminate subjectivity and bias in basketball analytics, it is important to understand the reasons why each player has a different value. Perhaps the simplest reason is that no two players have the same physical makeup or approach to the game. However, certain forms of shots and shot approaches generally have higher levels of success and the number of times a player attempts the high-success shots can help indicate his commitment to the team's success and understanding of the game of basketball in general. Additionally, a given player's rate of success on more difficult shots can show his natural ability to score in ways in which other players may not be able (Erčulj and Štrumbelj, 2015). While a player's individual skills can easily be misrepresented by the public's perception, his impact on his team is perhaps even more susceptible to misperception.

Flaws with the eye test. A coach may overutilize a player based on more easily discernible skills. If a player, due to his style of play and physical build, has a natural knack for hitting a certain shot or making a certain pass that would normally be considered difficult for most players, that player may receive extra playing time because the coach could clearly see this positive aspect of the player's game. Meanwhile, another player may have less success with the difficult shot or pass, and, thus, appear to rate lower on the so-called "eye test," but he might also have an exceptional ability to trick defenders with a pass and leave his teammates wide open to take easy shots, thus reducing the level of difficulty for the entire team. This second player may be more valuable to the team as a whole than the first player, despite his perceived lesser personal ability, because he so greatly improves the entire team's chances of successfully making shots. This would be a clear example of bias toward certain abilities, and this bias found in the "eye test" is not inherently eliminated by all forms of statistical analysis. Especially in the simplest forms of basketball statistics, such as points, rebounds, and assists on a per game basis, individual players can experience inflation or deflation of statistics relative to their total value to the team. For example, some players happen to score high numbers of points, not due to their own ability to score, but rather because they have teammates that happen to be very proficient passers and a well-coached team that successfully creates openings for that player. Thus, to judge the player's value to the team, or even value as a scorer, solely on his points scored per game would be an inaccurate valuation, as the statistics do not necessarily account for the surrounding factors in the player's performance.

Inherent bias. One of the clearest distinctions among NBA players is that of All-Star. An All-Star, or player who has been selected to the All-Star Game, will be considered among the very best in the sport for the season in which he is selected. Thus, finding the statistical traits common among All-Stars can help quantify the differences between players of varying skill levels. Additionally, by finding how players at different positions are clustered in regard to certain statistics, it is possible to better compare them and find the best players at each position (Sampaio, Janeira, Ibáñez, & Lorenzo, 2006). All-Star Games as a measure of player value are most certainly imperfect, however, as the players' selections are partially based on voting by fans, who may not be qualified to assess a player's true ability or value. An All-Star, however useful as a distinction, is still, to an extent, a perception-based concept, meaning a more concrete differentiation would be useful.

Early Statistics

The first statistics to become commonplace in basketball were the simplest ones, such as points and rebounds, often represented through per-game averages. These would be the type of counting statistics that are represented today in newspaper box scores. While the average fan surely appreciates the simplicity and easy comparability of these box score values, they are, by far, the most heavily biased pieces of data in basketball analytics. They do not account for any extenuating circumstances or influencing factors, such as teammate or opponent ability. Rather, they simply count totals and divide by the number of games played. Thus, these statistics were designed to thrill fans and name league leaders, not effectively analyze a player's overall value, and must be treated as such by analysts.

Advanced Statistics

An important part of determining the greatest players of all time is being able to determine which statistics best indicate the value of a player to a team. By dividing players by position and finding which statistics most often correlated to a player's team being victorious, it is possible to show that certain statistics can best discriminate between a player on a winning team and a player on a losing team. It is also possible to show which positions best exemplify the power of these statistics, perhaps making those positions more vital to a team's success (Escalante, Saavedra, & Garcí-Hermoso, 2010).

Team statistics. Over time, basketball statistics have grown more effectively descriptive and detailed. One such advancement is a greater dedication to tracking statistics for the team as a whole when a player is on the court, as this gives a greater representation of the player's actual impact on the team's success. The best-known example of this method of analysis is tracking a player's plus/minus ratings. The plus/minus rating, often averaged by game or per 36 minutes (a common amount of playing time in a single game for a player in a starting lineup), gives the difference in points between the player's team and their opponent for when that player is actively on the court and in the game. As the true goal for a player is to help his team win the game by scoring more than the other team, a plus/minus rating shows how effectively a player accomplishes that goal and, thus, increases his team's chances of winning.

Individual player tracking. Additionally, analysts have increasingly broken players' performances down into multiple categories. For example, by separating all the shots a player takes by the area on the court from which the shot was taken, it is possible to determine which areas and types of shots most efficiently utilize a player's unique skills. This can be used both to coach a player how to play on offense or how to guard another player while on defense. With each passing year, new statistics meant to further compartmentalize the analysis of the game are introduced and placed into regular use.

Current Industry Standards for Individual Player Analysis

Specific Event Trackers

Effective field goal percentage. The current landscape of NBA statistics greatly emphasizes precise breakdowns of game events into specific categories, when tracking both teams and individual players. One such example is effective field goal percentage, or $eFG\%$. Unlike a player's field goal percentage, which simply divides the number of made field goals by the number of attempted field goals, effective field goal percentage adds one half of made three-point field goals to the numerator. Thus, if *FGM* represents field goals made, FGM_3 represents three-point field goals made, and *FGA* represents field goals attempted, then field goal percentage, or $FG\%$, is $FG\% = \frac{FGM}{FGA}$, and effective field goal percentage, or $eFG\%$, is

$$eFG\% = \frac{FGM + 0.5(FGM_3)}{FGA}. \tag{1}$$

This gives credit to players who may sacrifice their own field goal percentage in order to attempt the more difficult, yet more rewarding, three-point shots more often.

Percentage of field goals assisted. Another commonly used statistic that tracks a highly specific part of a player's game is percentage of field goals assisted, or $FGM_{\%AST}$. This measures the amount of a player's made field goals that occurred because of a teammate's assist. If FGM_{AST} represents field goals made on which a teammate recorded an assist and FGM represents field goals made, then

$$FGM_{\%AST} = \frac{FGM_{AST}}{FGM}. \tag{2}$$

This enables analysts to determine both if a player is good enough to make shots without the constant help of his teammates and if a player is constantly choosing to do everything himself, which may, perhaps, be to the detriment of the team as a whole if his chances of hitting the shot are not reasonably high. Through these and myriad other modern statistics, teams can more effectively determine the tendencies of each individual player and more accurately assess their value to the team's success.

Random walk nature of scoring.

It can be argued that the points scored in a basketball game can be approximated by a random walk model, using probabilistic concepts such as the Poisson process and exponential distribution. Such a model could be used as a comparison to demonstrate a player's ability to either overperform or underperform the projected expectations. The random walk concept argues that scoring in basketball is a truly memoryless process, as opposed to a process with long-time correlation. The applicability of random walk theory discredits the concept of “hot” and “cold” streaks, saying rather that a player has a far more fixed probability of hitting any given shot and the memoryless-ness of the process allows hits and misses to occasionally come in bursts (Gabel & Redner, 2012).

Overall Performance Analysis

Player impact estimate. There are several more recent attempts to analyze players based on their overall value to their teams. This is, of course, very difficult to measure, as there are so many different statistics focused on specific measurements, and to combine them into a broader metric adds layers of complication. While very few metrics have been created that successfully encompass individual players’ complete performances, there are some, like *PIE* and *PER*, that do indicate a player’s complete value better than simply determining points per game.

The Player Impact Estimate, or *PIE*, attempts to factor as many different statistics into evaluating a player’s performance as possible as a percentage of the total number of those same statistics in the game across all players. Thus, using the legend in Table 1 and a “*Gm*” subscript to denote a statistic representing the total number for all players in the game, the equation is:

$$PIE = \frac{PT+FGM+FTM-FGA-FTA+DRB+\frac{ORB}{2}+AST+STL+\frac{BLK}{2}-PF-TO}{PT_{Gm}+FGM_{Gm}+FTM_{Gm}-FGA_{Gm}-FTA_{Gm}+DRB_{Gm}+\frac{ORB_{Gm}}{2}+AST_{Gm}+STL_{Gm}+\frac{BLK_{Gm}}{2}-PF_{Gm}-TO_{Gm}} \tag{3}$$

Though this formula still fails to account for the so-called intangible, or nontrackable, pieces of a player’s performance, it certainly considers a large number of aspects of ways a player’s measured performance relates to his team’s success as a whole.

Table 1. *List of abbreviations for statistics.*

Abbreviation	Statistic
AST	Assists
BLK	Blocks
DRB	Defensive Rebounds
FGA	Field Goals Attempted
FGM	Field Goals Made
FTA	Free Throws Attempted
FTM	Free Throws Made
ORB	Offensive Rebounds
PF	Personal Fouls
PT	Points Scored
STL	Steals
TO	Turnovers

Player efficiency rating. Another well-known example of a comprehensive statistic, perhaps the most commonly used and best known in basketball analytics in the modern era, is the Player Efficiency Rating, or *PER*. *PER* is far more complex than *PIE* and is designed to adjust for average performances across the entire league, allowing for comparisons between players on different teams and even in different eras. The average score for *PER* is 15.00 (Calculating PER, 2018), while the all-time career leader is Michael Jordan at 27.91. As *PER* accounts for both league and team averages and their impacts on a player's counting stats, the calculation is rather lengthy. The formula for unadjusted *PER*, or *uPER*, using many of the terms from Table 1, is

$$\begin{aligned} uPER = & \frac{FGM_3 + \frac{2 \cdot AST}{3} + \left(2 - \left(\frac{\text{factor} \cdot AST_{team}}{FG_{team}}\right)\right) FGM + \frac{FTM \cdot \left(2 - \frac{AST_{team}}{3 \cdot FGM_{team}}\right)}{2} - VOP(TO + DRB\%(FGA - FGM))}{MIN} \\ & - \frac{VOP((0.1936 + 0.2464 \cdot DRB\%) \cdot (FTA - FTM) - (1 - DRB\%) \cdot (RB - ORB))}{MIN} \\ & - \frac{VOP(DRB\% \cdot ORB - STL - DRB\% \cdot BLK) + PF \cdot \left(\frac{FTM_{lg} \cdot (1 - 0.44 \cdot VOP)}{PF_{lg}}\right)}{MIN}, \end{aligned} \tag{4}$$

where "*team*" denotes that the statistic is the average of all the players on the team and "*lg*" denotes that the statistic is the average of all the players in the league for the previous season, and factor, *VOP* (or value of possession), and *DRB%* (or defensive rebound percentage) each require formulas to calculate individually. These formulas are as follows:

$$\text{factor} = \frac{2}{3} - \frac{AST_{lg} \cdot FTM_{lg}}{4 \cdot (FGM_{lg})^2} \tag{5}$$

$$VOP = \frac{PT_{lg}}{FGA_{lg} - ORB_{lg} + TO_{lg} + 0.44 \cdot FTA_{lg}} \tag{6}$$

$$DRB\% = \frac{DRB}{RB}. \tag{7}$$

Now it is possible to adjust the metric for league and team averages using the following formula:

$$PER = \frac{15 \cdot uPER \cdot \text{Pace}}{\text{Pace}_{team}}, \tag{8}$$

where Pace is an estimate of the number of possessions by a team per 48 minutes played, calculated using the following:

$$\text{Pace} = 48 \cdot \frac{(Poss_{team} + Poss_{opp})}{2 \cdot \frac{MIN_{team}}{5}}, \tag{9}$$

with *Poss* equal to number of possessions and the "*opp*" subscript representing the opposing team's totals.

Thus, *PER*, while a highly sophisticated and advanced metric, requires a large amount of time and computational power to calculate for even a single player and still only truly credits players who successfully achieve the tangible counting statistics that show up in the formula. Despite the many advantages of the existing measures of analysis, they all fail to truly account for a player's value with respect to the whole team, instead substituting the simple statistics a player accumulates.

Original Research

New Formula

Justification for chosen data types. At its essence, the art of quantitatively determining anything in sports not explicitly given by the final score of a game is imperfect at best. This most certainly applies to finding the greatest NBA player of all time, as the greatest player would give his team the greatest probability of winning. Using techniques most commonly found among Vegas betting books and fantasy sports fanatics, it is possible to estimate, if not quite determine, the probabilities of a player's success in any given part of the game of basketball. Extending this to the impact the player has on the team can give a clear picture of the greatest player of all time (Winston, 2012).

Because each position on a basketball team is typically manned by a player with a unique skill set, the quantifiable statistics produced by each position have the potential to vary greatly. A discriminant analysis of each position in comparison to the statistical outputs of other positions allows statisticians to better adjust more advanced metrics to account for the specific value each player provides. In particular, centers and guards display the difference in statistical output, as they are responsible for different aspects of the game. Thus, it is necessary to find the statistics that favor one position over another and properly adjust for the difference (Sampaio et al., 2015).

The truest, most accurate analysis of individual players would, in some way, capture the myriad incalculable ways a player impacts his team that are not recorded by a specific statistic. This cannot be accomplished merely by considering more statistics. While it is possible to infinitely break down a player's individual scoring, it is impossible to determine the full effect that player has on his four teammates on the court. Perhaps his renowned shooting ability distracts an extra defender, leaving a teammate even the slightest bit more open to take an easier shot. It may be that a player has a low total of steals, but also very rarely allows passes to open opponents because his defensive skill dissuades opponents from taking the risks necessary to create openings. These types of impacts are not recorded in any way aside from the final score. Thus, the best possible individual analysis would be determined, not by a conglomeration of individual statistics, but rather by finding a player's usage by his team and the impact of the minutes he plays on the game's final score.

The use of network analysis accounts for a greater number of variables and greater lack of linearity in data sets than traditional regression techniques. At its core, network analysis in basketball statistics identifies bipartite graphs with two distinctive types of nodes – units (or lineups) and players (Skinner & Guy, 2015). A matrix can then be created using the degrees of incidence, or the number of connections, between the nodes. For player P_i and unit U_j and given matrix W, $W_{i,j}$ is equal to the degrees of incidence between P_i and U_j. If there are *m* players and *n* units and $\text{In}\left(P_i, U_j\right)$ is equal to the degrees of incidence between a player P_i and a unit U_j, then the incidence matrix, W, is:

$$W = \begin{bmatrix} \text{In}(P_1,U_1) & \text{In}(P_1,U_2) & \dots & \text{In}(P_1,U_n) \\ \text{In}(P_2,U_1) & \text{In}(P_2,U_2) & \dots & \text{In}(P_2,U_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{In}(P_m,U_1) & \text{In}(P_m,U_2) & \dots & \text{In}(P_m,U_n) \end{bmatrix}. \tag{10}$$

By multiplying this matrix by its transpose, we are given $WW^T = A$ with

$$A = \begin{bmatrix} \sum_{i=1}^n \text{In}(P_1,U_i)\cdot(U_i,P_1) & \sum_{i=1}^n \text{In}(P_1,U_i)\cdot(U_i,P_2) & \dots & \sum_{i=1}^n \text{In}(P_1,U_i)\cdot(U_i,P_m) \\ \sum_{i=1}^n \text{In}(P_2,U_i)\cdot(U_i,P_1) & \sum_{i=1}^n \text{In}(P_2,U_i)\cdot(U_i,P_2) & \dots & \sum_{i=1}^n \text{In}(P_2,U_i)\cdot(U_i,P_m) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n \text{In}(P_m,U_i)\cdot(U_i,P_1) & \sum_{i=1}^n \text{In}(P_m,U_i)\cdot(U_i,P_2) & \dots & \sum_{i=1}^n \text{In}(P_m,U_i)\cdot(U_i,P_m) \end{bmatrix}. \tag{11}$$

This matrix, *A*, represents a unimodal network, or a network where each single player/single unit combination creates a unique entry in the matrix. Thus, the network can be evaluated in terms of a single variable instead of multiple variables, as would have been the case before. In fact, this matrix encompasses all of a single player’s related units and allows for analysis strictly between players based on their involvement in different units. The edges, or connections, between players will be weighted based on the efficiency ratings of the units in which both players are involved. This enables us to find a centrality value for each player indicating the number of high efficiency units in which the player is involved. As a player’s centrality value could be inflated due to the number of other high efficiency players in the neighborhood of, or connected to, that player, an additional rating, known as a p-score, is created that indicates the level of artificial inflation in a player’s centrality caused by the players around him (Piette, 2011).

Utilizing these two metrics in tandem gives an accurate representation of a player's value to his team relative to the player's around him and with respect to players on other teams playing in their own units. By finding an appropriate method for combining the two metrics, it will be possible to create a single value for each player that displays the player's total value. This value can be used to compare any two players within a given era of play.

Likely the most consistent form of measurement between both teams and players across different eras and styles of play are the statistics built around a per-possession metric. Because any two teams in a given game are guaranteed to have a difference in number of possessions of at most two and each possession has a restricted number of outcomes, it is far easier to accurately compare performances. A player may score far more points per game than another player, but may be equal in points per-possession, which is more indicative of the overall impact by that player on the team. The resulting discrepancies can be accounted for by measuring both players and teams by their per-possession productivity, using both descriptive and mathematical explanations (Kubatko, Oliver, Pelton, & Rosenbaum, 2007). This concept of per-possession values and its relationship to efficiency can be utilized to give a weighted value to use in the previously discussed network analysis method, creating something of a level playing field for players used in different situations.

Explanation of final metric. For this specific study, the team efficiency aspect of the weighted value between two players, player i and player j, within the network, written as $E_{i,j}$, will be

$$E_{i,j} = \sqrt[4]{\frac{OPPP_{i,j}}{DPPP_{i,j}}}. \tag{12}$$

$OPPP_{i,j}$ and $DPPP_{i,j}$ stand for offensive points per possession and defensive points per possession while those players are both on the court, respectively. Note that if $OPPP_{i,j}$ is greater than $DPPP_{i,j}$, then $E_{i,j}$ is greater than 1. Inversely, if $DPPP_{i,j}$ is greater than $OPPP_{i,j}$, then $E_{i,j}$ is less than 1. As the objective would be to outscore the opposing team, a player should be rated higher if, while he is on the court, his team is successfully scoring more points per possession than their opponents are. The most effective way to quickly differentiate the weighted values between two players at this point would be to raise this determined factor to a power greater than 1. Thus, if $E_{i,j}$ is greater than 1, it will be raised even higher. Meanwhile, if $E_{i,j}$ is less than 1, it will be lowered further.

The exponent for this weighted value can be expressed mathematically, with LP_i for the number of lineups in which player *i* plays, PP_i for the number of players with which player *i* plays, and an "*avg*" subscript representing the average for all players, as the following:

$$X_{i,j} = \frac{LP_i + LP_j}{2 \cdot LP_{avg}} + \frac{PP_i + PP_j}{2 \cdot PP_{avg}}. \tag{13}$$

Thus, the more lineups in which a player plays and the more players with whom he plays, the greater his significance in the network representing the team and the higher his weighted centrality measure.

Each method of analysis has its specific values and purposes, but many are not designed, and thus should not be used, to analyze a player's overall value to his team. The special few that are, however, built to evaluate a player and rank him are unique in the statistics they emphasize and, therefore, experience occasional differences in the ranking order of players.

This research will aim not only to discover and discuss the biases inherent in different methods of analysis and how best to eliminate them, but also to demonstrate an effective way to combine aspects of network analysis into single metrics to create an easily compared output for each player. The two primary aspects on which we will focus are the centrality score and the p-score. By finding a relationship between these two values and combining them, it will be much simpler to compare players based on their value to their team's on court play. Using this single score, it will be possible to analyze the best players of all time and provide strong evidence for which one truly deserves the title of greatest.

To find the compiled weighted score between two players, *i* and *j*, the two components mentioned previously, $X_{i,j}$ and $E_{i,j}$, are utilized, in addition to $MIN_{i,j}$, or the number of minutes the two players played on the court together. This gives us the following equation:

$$W_{i,j} = \left(\frac{MIN_{i,j}}{2} \cdot E_{i,j}^{X_{i,j}}\right)^{\frac{1}{2}} = \left(\frac{MIN_{i,j}}{2} \cdot \left(\sqrt[4]{\frac{OPPP_{i,j}}{DPPP_{i,j}}}\right)^{\frac{LP_i+LP_j}{2 \cdot LP_{avg}} + \frac{PP_i+PP_j}{2 \cdot PP_{avg}}}\right)^{\frac{1}{2}}. \quad (14)$$

This, however, is not the total centrality score for a player. To find that, the weighted scores between that player and each other player must be summed, divided by the number of different players with which that player has appeared in a game, designated PP_i for player *i*, and the number of games in which that player has appeared, designated GP_i. Written as an equation, for player *i* on a team with *n* players, the centrality score is

$$C_i = \frac{\sum_{j=1}^{i-1} W_{i,j} + \sum_{j=i+1}^{n} W_{i,j}}{GP_i \cdot PP_i}. \tag{15}$$

In theory, this score measures the raw basketball efficiency a team experiences given a certain player, denoted *i* in this case, is on the court. It is easily compared to the centrality scores for other players to determine which players experience the greatest portions of the team's success. However, this rating still leaves something to be desired, at least when attempting to stand alone.

Many players will experience centrality scores that appear to unfairly give a player too much or too little credit for the team's efficiency. For example, a superstar player may do the majority of the work for the team, while another, less talented player may happen to play approximately the same amount of time on the court as the superstar, and often at the same time as the team's star player. In this case, the lesser player would experience similar team efficiency ratings and have a centrality score not much different than that of the better player. Thus, it is important to find a way to adjust scores such that they more accurately reflect not only the team's success while a given player is on the court, but also account for the other players most often on the court with that player and the contributions, or lack thereof, they make to the team. The simplest way to do this is by looking at the average centrality scores of the players with which the given player has been on the court. The equation form of this is

$$P_i = \frac{C_i \cdot \sum_{j \in N_i} \frac{C_i}{C_j}}{PP_i}, \tag{16}$$

where N_i is the set of all players on the team except for player *i*. This adjusted centrality score can serve as valuable insight on its own, though it also contains much value when paired with the unadjusted centrality score from which it is derived.

As may likely be imagined, the difference between the adjusted and unadjusted centrality scores can help to indicate a given player's value to his team relative to the values of his teammates, as well as be used by coaches and analysts to determine lineup adjustments to best utilize each player. If a player's Adjusted Centrality Score is far higher than his unadjusted centrality score, he is most commonly playing alongside players of lesser quality and value to their team than himself. Likewise, if his adjusted centrality score is far lower than his unadjusted centrality score, he is most commonly playing alongside players of greater quality and value than himself. While the value of this information to coaches may be clear, as it enables them to adjust the team's lineup and utilize players to the most efficient extent, it is also highly valuable to analysts discussing the overall value of players across different time periods and styles of play. The most valuable players will not only have high adjusted and unadjusted centrality scores, they will have a large difference between the adjusted centrality score and the unadjusted centrality score.

Application of Formula

Composition and justification of player list. A full application of this set of metrics across the careers of each and every player in NBA history would be impractically difficult. To apply it to just a select few players considered to be among the greatest ever would prove rather lengthy and cumbersome. Thus, that task should be saved for another undertaking. Here, we will simply work to show the viability of the metric when analyzing basketball players by taking the ratings from a single game, Game 3 of the 2018 NBA Finals between the Cleveland Cavaliers and the Golden State

Warriors. The Golden State Warriors of recent years are considered one of the best assemblies of basketball talent in NBA history, including two former MVPs, Steph Curry and Kevin Durant. The Cleveland Cavaliers' star player is LeBron James, widely considered one of the two best players in the history of the NBA, alongside Michael Jordan. Additionally, this game in particular was highly competitive, with the score remaining close throughout the contest, and took place in one of the highest-pressure scenarios in the game of basketball. Thus, this game will serve as an excellent way to compare some of the candidates for greatest players of all-time.

First, we will look at the Net Ratings, or the team's point differential per 100 possessions while the given player is on the court, and the Player Impact Estimates for the given game for each team in order to establish a baseline, which shall later be used to show that the newly created network centrality score metrics is a viable option for basketball analytics and player value determination. Since a player with low numbers of minutes played is more likely to experience statistical outliers from game to game and is generally considered to be less talented than the players with the largest numbers of minutes played, we will restrict this study to only the players who were on the court for at least fifteen minutes of Game 3 of the 2018 Finals.

Resulting rankings. As can be seen in Table 2, Kevin Durant had by far the highest Player Impact Estimate, as well as the most minutes played, of any Golden State Warrior. While Kevin Durant's Net Rating is not the highest, only Andre Iguodala's is sufficiently different to attract notice, and this may likely be attributed to the low number of minutes Andre Iguodala played relative to Kevin Durant, leading to a single-game outlier. Meanwhile, Draymond Green had, by far, the lowest Net Rating for the Warriors, despite his far more respectable Player Impact Estimate and high number of minutes played. Thus, in this particular game, his play, while individually impressive, was below the average in terms of its effect on the team's success.

Table 2. *Golden State Warriors' rankings by existing statistics.*

Player Name	Net Rating	Player Impact Estimate (PIE)	Minutes Played
Kevin Durant	21.3	27.5	43.3
Shaun Livingston	12.0	11.0	17.4
Andre Iguodala	41.3	8.7	21.9
Draymond Green	1.9	8.6	40.4
Klay Thompson	22.2	4.3	40.6
Steph Curry	3.5	3.2	39.2

Note: Players Advanced. (n.d.). Retrieved from http://stats.nba.com/players/advanced/

It is interesting to note the players with the second, third, and fourth-most minutes played were the three lowest ranked players for the Golden State Warriors by Player Impact Estimate, as Player Impact Estimate does not adjust for minutes played. Next, we will look at the same statistics for the Cleveland Cavaliers players who were in the game for at least fifteen minutes.

While the consensus remains that LeBron James was the best player on this team, the small sample size of a single game allows for Kevin Love to pass him in these rankings, as shown in Table 3. However, the Cavaliers' data, in general, matches expectations regarding player values. While his Player Impact Estimate was not overwhelmingly impressive, Tristan Thompson had the highest Net Rating for the Cleveland Cavaliers in this game with a high number of minutes played. Inversely, Rodney Hood had the lowest Net Rating on the Cavaliers despite a relatively high Player Impact Estimate with a low number of minutes played.

Table 3. *Cleveland Cavaliers' rankings by existing statistics*

Player Name	Net Rating	Player Impact Estimate (PIE)	Minutes Played
Kevin Love	-6.0	18.4	31.2
LeBron James	-14.5	17.7	46.9
Rodney Hood	-26.7	15.7	25.6
Tristan Thompson	-4.1	5.9	33.8
JR Smith	-10.3	4.4	33.1
George Hill	-12.5	-0.4	27.3
Jeff Green	-23.3	-1.7	18.0

Note: Players Advanced. (n.d.). Retrieved from http://stats.nba.com/players/advanced/

As would be expected based on their previous records and accolades, players such as Kevin Durant and LeBron James were among the leaders for their respective teams. One interesting outlier in this data is Steph Curry, who was last on the Warriors among those who played at least fifteen minutes in Player Impact Estimate and second to last in Net Rating, despite having won two previous MVP awards and being considered among the best shooters in the history of professional basketball. This can be explained, however, by the fact that he was recently returned from injury and not yet fully recovered. In fact, his playing ability appeared severely hampered throughout the playoffs, including this game. Thus, at least based on this small sample size, both Net Rating and Player Impact Estimate appear to be reasonably acceptable ways to determine a player's value to his team. However, as Net Rating does not adjust for potential outliers who play fewer than average minutes per game and Player Impact Estimate is based on a player's individual counting statistics, it would seem likely that a network analysis model, such as the centrality score metrics, would possibly be a more accurate manner of player evaluation.

Having shown that existing statistics are not completely out of line with expectations, we shall now utilize the newly created Centrality Score to analyze the raw data from a different approach. Once again, we will initially consider only players who played at least fifteen minutes in the game, though we will acknowledge additional players due to the fact that the centrality scores account for number of minutes played to reduce outliers stemming from a low number of minutes played. As the number of minutes played does directly and positively affect a player's Centrality Scores, a player with a higher Centrality Score than a different player who spent more time on the court likely played significantly better during his time in the game to make up the difference. This is explored in Tables 4 and 5, for the Warriors and Cavaliers, respectively. The player evaluations for the Golden State Warriors are reasonably similar, whether using the Player Impact Estimate or Centrality Scores. In both cases, Kevin Durant is the number one player and Steph Curry ranks last, as would be predicted, both by their individual statistics and the statistics for the team as a whole while those players were in the game. While Draymond Green and Shaun Livingston are ranked lower by Centrality Scores than by Player Impact Estimate, their Player Impact Estimates were not in line with their Net Ratings. Centrality Scores are more closely related to Net Ratings than

Player Impact Estimates, which would account for this difference. Andre Iguodala's exceptionally high Net Rating contributed greatly to his high Unadjusted Centrality Score, but as Klay Thompson played far more minutes in the game, he experienced a greater difference between his adjusted and unadjusted scores due to the extra value he created not measured purely in points.

Table 4. *Golden State Warriors' rankings by Centrality Scores*

Player Name	Unadjusted Centrality Score	Adjusted Centrality Score	Minutes Played
Kevin Durant	7.764	11.961	43.3
Andre Iguodala	7.117	8.698	21.9
Klay Thompson	6.858	8.856	40.6
Shaun Livingston	6.324	7.921	17.4
Draymond Green	4.107	3.702	40.4
Steph Curry	3.978	2.989	39.2

Table 5. *Cleveland Cavaliers' rankings by Centrality Scores*

Player Name	Unadjusted Centrality Score	Adjusted Centrality Score	Minutes Played
Tristan Thompson	3.992	6.806	33.8
Kevin Love	3.417	4.358	31.2
LeBron James	3.061	5.106	46.9
George Hill	2.860	3.386	27.3
JR Smith	2.618	2.632	33.1
Jeff Green	1.669	1.138	18.0
Rodney Hood	1.391	0.742	25.6

The Cleveland Cavaliers' Centrality Scores do not align with their Player Impact Estimates as well as the Golden State Warriors', but many of the differences can still be accounted for through simple reasoning. For example, Tristan Thompson is ranked three places higher by Unadjusted and Adjusted Centrality Scores than by Player Impact Estimate. However, he has the best Net Rating among the Cavaliers' players who played for at least fifteen minutes in the game and the second most minutes played, behind only LeBron James. These two factors, which are considered much more strongly in Centrality Scores than in Player Impact Estimate, are likely more than enough to create that difference between the rankings. Similarly, Rodney Hood, who had the lowest Net Rating among the Cavaliers' players with at least fifteen minutes played and the second-lowest number of minutes played, is ranked four places lower by Unadjusted Centrality Score than by Player Impact Score, as, despite his lofty individual statistics, the team was less successful than average while he was on the court. Another interesting aspect to these rankings is the much greater difference between LeBron James's Adjusted Centrality Score and Unadjusted Centrality Score than between Kevin Love's Adjusted Centrality Score and Unadjusted Centrality Score. In fact, LeBron James experiences a 66.8% increase, while Kevin Love only experiences a 27.5% increase. While, for this specific game, the Cavaliers typically performed better on average with Kevin Love on the court than they did with LeBron James, a deeper investigation of the data indicates that LeBron James played with weaker players more often than Kevin Love did. Those weaker players would drag down LeBron James's score by lessening the team's overall efficiency while he is on the court. Thus, the Adjusted Centrality Score recognizes the weakness of the players with whom LeBron most often played and gives him the credit for the success of those lineups. Meanwhile, the much lower increase in Kevin Love's score indicates that much of the team's success while he was on the court was due to his teammates' abilities nearly as much as his own. Simply put, LeBron James was asked to do far more with far less than Kevin Love was, and the Adjusted Centrality Scores reflect this. While there are certainly noticeable differences between the Player Impact Estimates and the Centrality Scores for the Cleveland Cavaliers, these differences appear sufficiently accounted for by the differences in team performance independent of players' individual statistics.

Final Results

Comparison to Existing Methods

As shown through the previous four tables, Centrality Scores compare favorably to other player evaluation analytics, such as Net Rating and Player Impact Estimate. In fact, the differences that do occur can be explained as potential failures on the part of the earlier systems to fully capture every aspect of a player's value to his team on the court. Additionally, because Centrality Scores account for minutes played and the number of utilized lineup combinations in which a player has appeared to reduce outliers, players who, perhaps, did not play enough minutes to clear an arbitrarily established threshold can still be compared to the players who did reach the required number of minutes. For example, Kyle Korver only played 10.8 minutes for the Cleveland Cavaliers, but had the sixth highest Unadjusted Centrality Score and Adjusted Centrality Score on the team for the given game. This is due in large part to the Cavaliers' Defensive Rating, or points allowed per 100 possessions. While Kyle Korver was on the court, the Cavaliers'

Defensive Rating was well below their average for the game. Additionally, Korver had the second highest Net Rating on the Cavaliers for the given game. By allowing the Centrality Scores to account for minutes played, we can see that Kyle Korver managed to contribute more to his team in his limited time in the game than multiple other players who played more minutes were able to contribute.

While this study is not comprehensive enough to determine the best player on each of the two teams considered, much less the greatest NBA player of all time, it does serve strongly as evidence that the Centrality Scores system of player analysis successfully ranks players based on their overall value to their team's success and compares favorably to other methods of analysis, such as Player Impact Estimate and Net Rating. While most other methods of player analysis evaluate a player using his individual statistics, or in the case of Net Rating, a team average with no regard to potential outliers, the Centrality Scores method utilizes a team's success related to the player with various safeguards built into the calculations to reduce the frequency of outlying data points.

Exploration of Potential Biases

In addition to providing a potential standard for player greatness based on value to the team as a whole, the Centrality Scores system can be utilized by coaches to compare players on a single team and determine the most successful lineup combinations and playing rotations for each player. By being able to track the value of a player relative to the other players on the court with him at a given time, coaches and analysts can tell if a player is truly benefitting the team when he plays or, rather, benefitting from the rest of the team for which he plays. To use an example from the earlier tables, Tristan Thompson of the Cavaliers had the best Unadjusted Centrality Score and Adjusted Centrality Score on his team for the game. Additionally, his Adjusted Centrality Score was 70.5% higher than his Unadjusted Centrality Score, meaning that he was even farther above the average player than his Unadjusted Centrality Score would indicate. Thus, it can be concluded that Tristan Thompson not only played in the best lineups that the Cavaliers had for this game, he was a primary reason those lineups were the most successful.

While arguments will always persist, whether amongst coaching staffs planning their strategies or fans debating the iconic status of past great players and their value, the Centrality Score system for analyzing player values establishes a baseline for comparing and contrasting players with their teammates. This is especially valuable to the debate over the greatest player of all time. While Michael Jordan has never played against LeBron James, it can now be determined whether Michael Jordan was more valuable to the Chicago Bulls than LeBron James to the Cleveland Cavaliers or Miami Heat. At that point, by comparing the efficiencies of their respective teams to the average efficiencies across the NBA during the years they played, it would be possible to figure out which of the two players provided the most overall value while on the court, making them the greatest of all time.

Coaching basketball is a very difficult job, with innumerable nuances and unforeseeable factors influencing the outcomes of every decision. The use of statistics and analytics does not completely solve this problem, but it enables coaches to look back at the results of their coaching and search for trends that may hold predictive value. Because Centrality Scores are designed to measure a player’s value in relation to a team’s overall success, they can be used to show both right and wrong decisions a coach made while keeping track of the various factors in those outcomes, which can then be further analyzed to illuminate the small tactical changes that can be made to improve both the individual players and the team as a whole. While coaches’ choices certainly impact a player’s Centrality Scores, these metrics offer a relatively objective measure of a player’s value for those in basketball seeking excellence.

References

Albert, J., Bennett, J., & Cochran, J. J. (2008). *Anthology of statistics in sports*. Philadelphia, PA: Society for Industrial & Applied Mathematics.

Albert, J., Glickman, M. E., Swartz, T. B., & Koning, R. H. (2017). *Handbook of statistical methods and analyses in sports*. Boca Raton: CRC Press.

Batchelor, B. (2013). *Basketball in America: From the playgrounds to Jordan's game and beyond*. London: Routledge.

Berri, D. J., & Eschker, E. (2005). Performance when it counts? The myth of the prime time performer in professional basketball. *Journal of Economic Issues, 39*(3), 798-807. doi:10.1080/00213624.2005.11506847

Bommel, M. V., & Bornn, L. (2017). Adjusting for scorekeeper bias in NBA box scores. *Data Mining and Knowledge Discovery, 31*(6), 1622-1642. doi:10.1007/s10618-017-0497-y

Calculating PER. (n.d.). Retrieved from http://www.basketball-reference.com/about/per.html

Erčulj, F., & Štrumbelj, E. (2015). Basketball shot types and shot success in different levels of competitive basketball. *Plos One, 10*(6). doi:10.1371/journal.pone.0128885

Escalante, Y., Saavedra, J. M., & Garcí-Hermoso, A. (2010). Game-related statistics in basketball by player position and final game score differences in European basketball championship 2007. *Fitness & Performance Journal, 9*(2), 50-56.

Fearnhead, P., & Taylor, B. M. (2011). On estimating the ability of NBA players. *Journal of Quantitative Analysis in Sports, 7*(3). doi:10.2202/1559-0410.1298

Gabel, A., & Redner, S. (2012). Random walk picture of basketball scoring. *Journal of Quantitative Analysis in Sports, 8*(1). doi:10.1515/1559-0410.1416

Gomez, M. A. (2008). Game-related statistics that discriminated winning and losing teams from the Spanish men's professional basketball teams. *Collegium Antropologicum, 32*(2), 451-456.

Ibáñez, S. J., Sampaio, J., Feu, S., Lorenzo, A., Gómez, M. A., & Ortega, E. (2008). Basketball game-related statistics that discriminate between teams' season-long success. *European Journal of Sport Science*, *8*(6), 369-372. doi:10.1080/17461390802261470

Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports, 3*(3). doi:10.2202/1559-0410.1070

Mcgarry, T. (2015). *Routledge handbook of sports performance analysis*. London: Routledge.

Okamoto, D. M. (2011). Stratified odds ratios for evaluating NBA players based on their plus/minus statistics. *Journal of Quantitative Analysis in Sports, 7*(2). doi:10.2202/1559-0410.1320

Otten, M. P., & Miller, T. J. (2015). A balanced team wins championships: 66 years of data from the National Basketball Association and the National Football League. *Perceptual and Motor Skills, 121*(3), 654-665. doi:10.2466/30.26.pms.121c25x4

Piette, S., Pham, L., & Anand, S. (2011, March 4). *Evaluating basketball player performance via statistical network modeling*. Lecture presented at MIT Sloan Sports Analytics Conference in Massachusetts Institute of Technology, Boston, MA.

Players Advanced. (n.d.). Retrieved from http://stats.nba.com/players/advanced/

Sampaio, J., Janeira, M., Ibáñez, S., & Lorenzo, A. (2006). Discriminant analysis of game-related statistics between basketball guards, forwards and centres in three professional leagues. *European Journal of Sport Science*, *6*(3), 173-178. doi:10.1080/17461390600676200

Sampaio, J., Mcgarry, T., Calleja-González, J., Sáiz, S. J., Alcázar, X. S., & Balciunas, M. (2015). Exploring game performance in the National Basketball Association using player tracking data. *Plos One*, *10*(7). doi:10.1371/journal.pone.0132894

Severini, T. A. (2015). *Analytic methods in sports: Using mathematics and statistics to understand data from baseball, football, basketball, and other sports*. Boca Raton: CRC Press.

Skinner, B., & Guy, S. J. (2015). A method for using player tracking data in basketball to learn player skills and predict team performance. *Plos One*, *10*(9). doi:10.1371/journal.pone.0136393

Winston, W. L. (2012). *Mathletics: How gamblers, managers, and sports enthusiasts use mathematics in baseball, basketball, and football*. Princeton: Princeton University Press.