Predicting Potential NBA Performance From Collegiate Statistical Data

A Thesis Presented to the School Of Business, Quinnipiac University

In Partial Fulfillment of the Requirements for the Degree Master In Business Administration

By

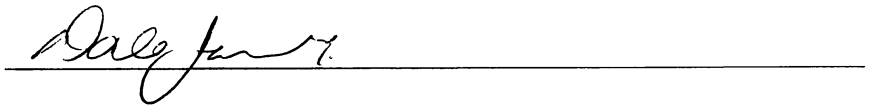
Cornelius Olufemi Thomas Jr.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

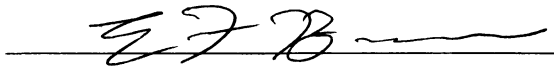
December 2005
Approval Page

Predicting Potential NBA Performance From Collegiate Statistical Data

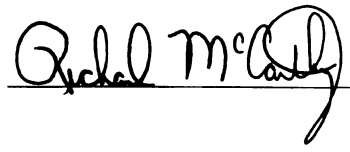
This thesis is approved as an independent investigation by a candidate for the degree of Master Of Business Administration and is acceptable as meeting the thesis requirements for this degree, but without implying that the conclusions reached by the candidate are necessarily the conclusions of the school.

Thesis Advisor 

Professor Dale Jasinski
Associate Professor Of Management
Quinnipiac University

Reader 

Professor Eric J. Brunner
Associate Professor Of Economics
Quinnipiac University

Director, MBA Program 

Professor Richard McCarthy
Quinnipiac University

Acknowledgements

First, I want to thank God for giving me the opportunities he has given me throughout my educational career. Thank you, God for giving me strength through the toughest of times.

I want to thank the most wonderful woman in the world, my lovely wife-to-be, Sherease Reese. Without her, I would never have made it through this journey acquiring my MBA. She always encouraged me to pursue my goals and do what I felt in my heart was best to do. For this, I will forever be grateful. I Love You.

I want to thank my parents, Cornelius O. Thomas Sr. and Abosede Akingbade, for providing the strong educational foundation for me to become the professional that I desire to be. Without their financial support, I would not have been able to fund pursuing my MBA. Thank You, Mom and Pop.

I want to thank Ben And Sharon Thompson for their continued support and encouragement throughout graduate school. You both have afforded e the opportunity to have a second "Mom and Dad" How lucky I am to have that luxury!

I want to express my deepest gratitude to the following individuals for their help and assistance in the completion of this thesis study:

I want to thank Eric Brunner, for taking a completely new student and teaching him the ins and outs of statistics. Your course gave me the confidence to continue to push forward and believe I had the capabilities to accomplish my thesis goals

I want to thank Duncan Fletcher, for listening, guiding, and putting up with my desire to work as hard as possible at all times, even when there was little need to. It has been the greatest pleasure working with you and PATI. You are destined for greatness in the NHL.

I want to thank Professors Angela Walker, Carrie Bulger, and Penny Leisring. It was the greatest honor to get to work with you three outstanding ladies and professionals. You all exemplify amazing professional women with great minds and I know that your accomplishments in your profession will be of the highest levels in the future.

I want to thank Professor Christopher Ball. I've watched you and admired your style and knowledge in economics. I've appreciated the conversations we've had and your interest in my capabilities and pursuits. Thanks for the "movie" of your life. You win the academy award.

I want to thank Professor Henry Adobor, for our lunches, and talks about business strategy and how to bring it to the NBA in the future. Thank you for being a wonderful example of a mentor and guiding me along this journey called graduate school.

I want to thank Professor Jasinksi for the opportunities that he has given me to serve as his graduate assistant, to perform my thesis, and to be apart of something bigger than me, PATI. Thanks

I want to also thank Professor Patrice Luoma for accepting me into the MBA program despite my youth and always being a sounding board for my career decisions.

I want to thank all my NBA contacts, Matt Simon, Isiah Thomas, Garry St. Jean, Jeff Nix, Steve Mills, Milt Newton, Dick and Scott McGuire, and Bobby Lewis. All of you have provided me with the inspiration to continue to try to make it in the League. Continue your great accomplishments and I plan on being there soon.

Table of Contents

List Of Tables

Predicting Potential NBA Performance From Collegiate Statistical Data

Cornelius O. Thomas Jr.
Quinnipiac University (2006)

Abstract

Purpose: This study seeks to find if there is a possible method of predicting potential performance for the NBA player from analyzing college statistics.

Design/Methodology: This study uses statistical data from the drafts from 1999-2003. The statistical categories of information recorded will be from per game averages of minutes, points, rebounds, assists, field-goal percentage, free-throw percentage, and 3-point percentage. Regression analysis will be used to analyze this data for trends or correlations. It will also be utilized to measure any vast differences or trends that occur on a position-by-position basis.

Findings: The results showed that there were significant t-stats for a number of statistical categories, though not all. After conducting a multiple variable regression analysis, the combination of NCAA PPG, TPG, SPG, 3PT FG%, College Class and FG% were shown to have the highest R-squared at .356. There was a negative relationship that occurred between NBA performance and college class.

Research Implications: There has been distinct interest in the athletic community for the development of methods by which to closely predict or project performance. Though the game has evolved significantly in the last few years, a 5 year parameter will allow for a relevant analysis of the development of a player throughout an average NBA playing career.

Introduction

The National Basketball Association is recognized as the highest level of professional basketball in the world. It consists of 30 teams that compete against one another in an 82-game season. This season normally begins during late October/early November and concludes during late April/early May. In its earliest years, the NBA was comprised solely of players who resided in the United States. Over the last 2 decades, the NBA integrated top players from international leagues to continue elevate the level of its play

Each year, the NBA selects its players through an event called the NBA draft held every June. During the draft, teams select players from all over the world to build and improve the talent level on their teams. The draft process in the NBA is compared to an "unraveling" of the labor market in the business world. Much of this sentiment stems from a belief that the draft is "a centralized matching system that places the top qualified athletes within organizations that have a distinct need for their services" (Groothis, Hill, and Perry, 2005). Every team's need for specific athletes varies according to the team's record and success during the season preceding the draft. The NBA conducts the draft in this manner due to the belief that this selection process most consistently allows for optimal attainment of personnel that can make a significant impact on team performance. The draft is constructed in a manner that suggests that there is a direct relationship between a high draft selection and positive performance. Unfortunately, after a brief survey of past drafts, it has become clearly evident that there is little proof to support the belief in this

direct relationship. In fact, after the first 3 selections there is little consistency in the relationship between high performance and draft position. High cost and risk of poor selection make successful drafting essential. To further familiarize you with the draft and its intricacies is a brief history of the draft.

The History Of The NBA Draft

The NBA Draft is an annual draft held including all of the 30 teams in the NBA where players are selected from a global pool of talent. Players are selected in two rounds and each team usually possesses a draft pick in each round. Teams may either trade their picks or stockpile draft picks by acquiring picks through trades or compensatory means. Players can be selected from colleges in the US, international leagues, and high schools in the US. At the present time, there are no definitive limitations as to those eligible for selection yet it is required that players must have either, 1) completed their senior year in high school or at least be 18 years of age before the year of the draft.

Between 1947-1965, territorial picks were issued to teams to help boost fan interest in cities of NBA teams. By this method, teams were also given the opportunity to forfeit their first round picks to select local talents that they projected would be able to contribute to the success of a NBA franchise. Over time, a new method of issuing draft picks was adopted where a coin flip would take place. The exact process was as follows:

1. A coin flip was conducted between the last place finishers in each of the two divisions to determine which team would open the draft.

2. The remaining teams would pick in inverse order based on their records in the prior season.

Though this process proved to be sufficient, it was flawed in that it motivated teams to "tank" the season as soon as they incurred a certain amount of losses. This strategy was adopted in order to maximize the possibility of landing a star talent by selecting early in the draft. Ultimately, as it became more evident that this strategy was being employed on a regular basis, a new draft process was employed. This process would be called the draft lottery.

In the draft lottery process, implemented in 1985, the draft lottery took into account a number of losing franchises and instead of assuring the two worst teams, the selection of the highest two draft picks, the top picks were placed into a lottery, particularly with the odds being increased of the worst team to select the first pick. By no means did it assure the worst team would be resurrected by a #1 pick and it diminished the incentive of teams to lose games purposely

In 1987, a more efficient procedural change was put into place where the lottery would determine the selection process for the worst teams only. The other team's selections would go by their records of the prior season. This helped diminish the possibility that a team with one of the worst three records in the league would

have high draft pick and diminish the possibility that a poor team would get a lower draft pick. This was done with a focus of preserving competition in the league.

Prior to 1989, all the drafts were conducted in a manner where players were selected until there were no more prospects. The draft rounds went from 21 to 10 to 7 by 1985. In 1989, it was decided by the players association and the NBA to limit the draft to just two rounds to allow undrafted players the opportunity to try out for any team.

In 1990, the NBA adopted a weighted draft lottery system where The team with the worst record during the regular season received 11 chances at the top pick (out of a total of 66), the second-worst team got 10 chances and the team with the best record among the non-playoff clubs got one chance. In 1993, the league again adjusted the lottery with hopes of increasing the likelihood of the worst team acquiring a top pick while simultaneously deterring purposeful losing by teams in season. The process was as follows:

1. 14 Ping-Pong balls numbered 1 through 14 are placed in a drum. There are 1,001 possible combinations when four balls are drawn out of 14, without regard to their order of selection.

2. Prior to the Lottery, 1,000 combinations are assigned to the Lottery teams based on their order of finish during the regular season.

3. Four balls are drawn to the top to determine a four-digit combination. The team that has been assigned that combination will receive the number one pick.

4. The four balls are placed back in the drum and the process is repeated to determine the number two and three picks.

This new system increased the chances of the team with the worst record drawing the first pick in the draft from 16.7 percent to 25 percent, while decreasing the chances of the team with the best record among lottery teams from 1.5 percent to 0.5 percent. From 1996 to today, the lottery has been increased to include 14 teams.

Introduction – (Continued)

In a time when performance evaluation is increasingly important to firms looking to maximize efficiency in operations, it is extremely vital to possess the capability to determine optimal methods for task accomplishment. Similarly to the business world, the sports world has identified that this search can either provide the foundation for either organizational successes or organizational failures. In an attempt to discern the best possible approach for personnel evaluation, sports organizations have utilized numerous technological resources that are available to today's business and public community. Sports organizations have began to record situations where statistics have been used to derive optimal lineups for maximum

beneficial performance in specific sports. Statistics have begun to provide a strong foundation for performance analysis in a number of sports arenas. In baseball, statistical analysis has become the motivation behind much of the decision making process in many front offices. Certain statistical performance levels can signify when players are performing at higher or lower levels. This is particularly important during the development process, where teams can use statistical data to decide when to promote, demote, or release players. The use of data works to simplify the evaluation and selection process.

This study is centered specifically on the sport of basketball. Basketball teams have become increasingly dependent on statistical analysis to make more sound decisions and investments on players. In basketball, it is vitally important to understand the significance of certain statistical accomplishments, particularly at the collegiate and professional levels where statistical analysis has become a requisite for the successful program/organization. Many important personnel decisions during a basketball season are dependent on statistical data compiled during the course of a season. Fans and evaluators alike have adopted general indicators of productivity (PPG, RPG,APG.....etc.) as the only justifiable resources from which to delineate player significance and performance.

The Goal Of The Thesis Draft Analysis

This thesis looks to analyze the selection method by using statistical data of players selected in the US over a 5-year period and examine a number of relevant

statistical categories to performance in NBA players. Upon examining the statistical data, data analysis will be conducted to develop a formula that will lead to a consistent method by which to measure positive performance of NBA players. After this formula is developed, the collegiate stats of players in the NBA will undergo a data analysis with the hopes of connecting successful or unsuccessful college performance to successful or unsuccessful performance in the NBA. It is my hope that this thesis will contribute to the depth of knowledge existing on data analysis of statistical data and performance evaluation in the NBA. Previously, most of the data analyses conducted on NBA statistics have not focused on the efficiency of the draft process. By mere examination of this process, it can be understood that any contributions or knowledge attained from this study will benefit those in the league in search of a better way to eliminate the risk associated with the selection process of NBA players.

The thesis will attempt to discover the most simplistic method that can be utilized to improve the efficiency of the selection process. With the understanding that the draft selection process is heavily based on statistical data, I believe there must be a possible method by which to use statistical data to predict potential positive professional performance of collegiate players.

Theory/Theoretical Implications

The Relevance of Statistical Analysis To The Evaluation Process

Statistical Analysis Rundown

In order to be successful in today's competitive environment, teams must utilize other methods of evaluation besides pure visual analysis. The naked eye has proved too inefficient to be solely relied on to select players. Coaches, general managers, and other front office personnel have searched for methods utilizing statistical data to improve the performance of their teams. Coaches have applied the use of statistical data as a supplement to their knowledge and decision-making process (Geitscher, 2005). Despite the significance impact that statistical data has had on improving the decision making process, there are those coaches and front office executives who believe more strongly in the strength of the scout's eye and understanding of the game. As the costs for acquiring players have increased, owners have chosen to employ individuals who can analyze data to avoid severe financial pitfalls that hurt their team's profitability.

During the early 1990's, the New York Knicks of the NBA served as pioneers in the statistical analysis field by utilizing an IBM program called Advanced Scout. This program acquired statistics from team personnel and analyzed statistics in an attempt to discover patterns that existed during the team's most successful performances. The video coordinator for the Knicks at the time, Robert Salmi, became an expert user of this program. This program marked one of the earliest documented uses of statistical data analysis by computer to influence performance. IBM created this program due to their desire to be at the forefront of data mining for sports performance. Before the creation of IBM's data mining resource, Advanced

Scout, MDS Qantel created Sports Pac, a program that provided capabilities to teams ranging from play-by-play analysis and scouting to daily medical treatment reports. The company had primarily offered the use of their program to businesses yet found a more interested partner in the sports world. The primary source of statistical sports data in the United States is STATS, Inc. Stats Inc. has dominated the statistical data acquisition market by developing the computer capabilities to track records and create interesting categories of statistics that improved the analysis of team performance.

At its most fundamental levels, statistical analysis is a form of data mining. Data mining is a method that takes statistical data that has been compiled and electronically analyzes it with a goal of attaining knowledge on a subject or issue. Data mining, in its nature, focuses primarily on the conversion of information into useful knowledge. Due to the fact information is ever-changing, methods of data mining will always be evolving. Businesses have made data mining in many of their operations as it has presented a way to improve services and capabilities. The sports arena is relatively new to data mining and available software to compute data efficiently is limited to date. Over time, the efficiency of this software will increase and become extremely powerful. Though data mining is relevant to a number of different professional arenas, our focus is on its relevance to sports. Data mining has become particularly relevant to NBA operations as data mining software like Advanced Scout, have been utilized regularly by NBA franchises.

Advanced Scout: The First Of Its Kind

Created in 1996, Advanced Scout was centered on the premise of making the coaching process ever more efficient. It is a reality that the human element in sports tends can, at times, have a negative effect on the consistency of performance. The goal of Advanced Scout was to eliminate the inconsistency that exists in coaching decisions of failure. It allowed coaches and assistant coaches to assess the effectiveness of certain decision that were being made. This product provided additional value to the fan by improving the quality of play and decision making.

The program is a piece of data mining software that functions in the following steps:

1. It collects data in connection with every aspect of the shooting, passing and rebounding process.

2. Each action is given a time code

3. At the end of each game the data is uploaded and stored where any team can request the information for its own analysis

4. Next it examines the data for consistency. It attempts to eliminate any possible errors that may accumulate during the data collection process

5. Upon the "cleaning" of the data, it is "translated for review by a coach. The data must be "translated" into a form that it is more connected to a coaching perspective. It puts the data into a play-sheet, a form commonly used by coaches in analysis of game situation.

To further facilitate the usefulness of the data, players are assigned roles. In addition to these steps, data and players are assigned role relationships which help to accurately assign and distribute weights that distinguish to the program the capabilities and roles of the players. This helps the program to correctly perceive the possible functions a player may undertake.

The program also gives users the opportunity to make inferences about various different trends that may be occurring in an opponent's game play. It allows individuals to query about any such trends and gain a vast understanding of the tendencies that its opponents may have. It connects trends in play to statistical categories which helps to connect the relevance of game play to statistical evaluation.

At the root of the program's capabilities is a concept called Attribute Focusing. It associates attributes with data sets to create consistencies to glean information. It takes subsets of data and looks for consistent correlation. When inconsistent correlations are found, they are marked because they happen to indicate tendencies that may be occurring in play that should be taken note of.

The information of relevance is connected and conveyed to users either by graph or text document. This helps to further facilitate the processing of intense data and information by coaches.

The Relevance Of The Advanced Scout Concept To The Thesis

Though the Advanced Scout program has a slightly different focus from the thesis being conducted, the process of data mining of statistics connects the two. The use of statistical basketball data requires that an understanding of the application of basketball statistics to data mining be understood. The Advanced Scout program represents an example of an innovative approach, adopted by the NBA's franchises, to maximize the decision-making process. Ultimately, the goal of this thesis falls along the same lines as the purpose of Advanced Scout. This thesis searches to make improvements in the processes of evaluation and decision-making that exists in the NBA by analyzing data.

Ultimately the Advanced Scout program was an attempt to improve managerial efficiency. The thesis proposed has the same goal. Like many other professional environments, the NBA has developed its share of inefficient management practices. Limited access to sufficient information analyses programs and methods have rendered general managers limited in the decision-making process. Of the 30 teams in the NBA, 26 are over the salary cap threshold and therefore are incurring luxury tax commitments.

Similar Studies Of Performance Measurement Through Statistical Analysis

Statistical Analysis, though commonly regarded for its ability to contribute to quality forecasting, is underutilized in the sporting arena. My thesis attempts to

contribute to the development and education of the public to the immense relevance of statistical evaluation methods in the accurate operation and construction of sporting teams. With a direct focus on the NBA, research work for this thesis has been limited to relevant measures of production through analysis of statistical information in the NBA. I believe that other additional research in other sports, besides the NBA, can be properly used to the benefit of my thesis and statistical analysis research.

An interesting NFL research effort showed that the uncertainty of the draft process can be tempered by the continual acquisition of information that may be relevant to the selection of a quality "employee" (draft pick) (Hendricks, DeBrock, and Koenker, 2003). Hendricks, DeBrock, and Koenker (2003: 860-867) employed previously used econometric models and relevant data were employed to accurately estimate possible risk aversion and option value associated with the draft selection process. These complex models involve the input of statistical data connected to productivity to estimate possible performance to fine tune the hiring process through the NFL Draft. These econometric models hold the concept of risk avoidance to be true under certain parameters that ensure definitive failure. Hendricks, DeBrock, and Koenker (2003: 884) reinforce that a measured risk may warrant certain selections during the "hiring process" despite existing levels of uncertainty.

Another document that examined the NBA showed that other studies have undertaken the goal of displaying the relevance and utilization of statistical data to measure performance. This study examined the MVP (Most Valuable Player) selection process currently utilized by the NBA (Berri, 1999). Despite the illogical

method of relying solely on media observations of productivity, the NBA has strongly continued to invest in this process as an accurate decider of value. References are made to the method employed by IBM to identify the MVP as it was in opposition to public opinion. Examinations occurred that investigated the true factors of productivity that most accurately consistent with productivity levels that warrant MVP consideration (Berri, 1999). It was identified that productivity and performance are most accurately measured by taking the marginal product of each player, as well as ensuring that there was a direct correlation between player point production vs. opponent point production. The study found that player value and productivity may be most closely correlated to a player's proficiency at shooting, rebounding, and avoiding turnovers (Berri, 1999).

Ultimately, the fact more studies are not undertaken with a relevance to the impact of statistical data show the need for more education on the effectiveness of data analysis. The hope is that my thesis can serve and contribute to the overall sporting statistical analysis community and increase awareness about the positive outcomes that can result form data research.

The Significance Of Sound Management

Improvement Of The Management Process

There is prevailing thought that when performance lags, the stimulant that can lead to an adverse result is change in leadership. Though many times this is the

case, without evaluating the true source of inefficiency in the production process, organizations short change themselves and incur an expensive hire with a marginal improvement in performance. The earlier mention *Advanced Scout* program offered a method of improving and adding to the evaluation of manager's efficiency yet it was met with reluctance. The thesis looks to follow a similar path as it searches to improve an area where historically there have been high percentages of failures and inconsistency. If management continues to reject innovative methods of improvement in the sports arena, there is high possibility that the quality of the sporting world will diminish at a rapid rate.

Consistently, organizations have disregarded the optimal value of their assets because of the organization-wide belief that as long as immediate rewards are present, the company's state is positive. This perspective is dangerous as it leads to complacency and lags in management proficiency. The difficulty in assessing managerial proficiency can be traced to the inability to control external factors contributing to performance. Normally, external factors have a distinct effect on the productivity of firms, it becomes difficult to distinguish to what degree productivity lags can be traced to management executives vs. environmental situations. Data Envelopment Analysis (DEA) is utilized to measure managerial productivity (Fizel and D'Itri, 1997). This method employs an approach where measurements are made between the differences in performance of managers versus results attained utilizing the consensus optimal decisions. There employment of this analytical method was utilized primarily in the field of basketball. This makes DEA an excellent example or guideline for this thesis to course its action.

What Influences Success In The NBA

To date, most of the studies regarding the success of players drafted focuses on the psychological factors that influence success. Studies conducted by researchers show that relationships do exist between psychological factors and on-court performance (Shoten, 2004). With most of the emphasis on NBA research being centered on the psychological field, a need exists for investigation into other factors that influence success among NBA players/draft picks. Statistical analysis can serve as that other investigative landscape that could possibly yield insight and information into what indicates and influences success in the NBA. (Geitscher, 2005) Past studies have shown that there has been little to no correlation between so many of these psychological factors often assessed by teams, yet strong relationships exist between statistical categories and performance success. High numbers in categories like, field goal percentage, defensive rebounding, assists per game, blocks per game, and points per game and success in standings, have shown a more consistent connection to successful performance. This shows that among introductory research, the relevance of statistical data and performance seems to be strong. Specific regression data showed that there was a positive relationship between FG%/ 3PFG% and winning percentage (Onwuegbuzie, 1998). Increases in winning percentages were shown to be closely connected to increases in field goal percentages (Onwuegbuzie, 1998). This connection, furthered reinforced through confidence intervals, showed that great validity exists in the analysis of statistical data and performance measurement in the NBA.

Hypothesis

The preceding discussion is the basis for the following hypothesis:

H_1 – There is a direct positive relationship between NCAA performance and NBA performance.

H_2 – There is a positive relationship between years in college/college class and length of NBA career.

Empirical Specification

Estimation Equation:

$$NBA\_Perf_i = \beta_0 + \beta_1 NCAA\_Perf_i + \beta_2 Z_1 + \varepsilon_1,$$

Description Of Estimation Procedure

Where NBA_Perf_i is a measure of player I's performance in the NBA, namely points per game, $NCAA\_Perf_i$ is a vector of variables that measure player I's performance while in the NCAA (e.g. points per game, rebounds per game, etc....), Z_1 is a vector

of personal characteristics such as player's height and weight and ε_i is a random disturbance term.

Expected Signs Of Coefficient

It is believed that there is a direct positive relationship between positive NCAA performance and positive NBA performance. It is also believed that there is a positive relationship between years in college and NBA career length.

Data

The data was obtained from a variety of sources mainly www.databasebasketball.com. This site serves as a highly regarded statistical site that has some of the most comprehensive statistical data on current and past NBA Players. Also utilized in the study was NBA.com, the NBA's mother website. There I was able to obtain statistical data on current and former players.

The key variables used were PPG, MPG, RPG, APG, SPG, TPG, FG%, ft%, and 3PT% at both the NCAA and NBA level. These variables serve as seasonal averages of player's statistical production. The total amount of each category is divided by the number of contests a player competes in. These categories have been used for almost 50 years to judge player productivity and performance. In this study, we will use NBA Career PPG as the main barometer for NBA productivity. We

assume this as it is widely regarded as the statistical category that is held in highest regard when rating or evaluating a player.

Categories/variable that were not utilized were A/TO ratio and per 48 minute production. The access to this data was limited and was not substantial enough to the final outcome of the regression analysis to warrant extensive research.

Analysis and Results

The basic summary statistics are as follows:

Summary Model Mean, Mode, Median Etc.....

Table I

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean		Std	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
HEIGHT	160	16	69	85	12630	78 94	27	3 425	11 732
WEIGHT	160	115	165	280	35405	221 28	2 12	26 814	718 996
Years In NBA	160	5	1	6	540	3 37	12	1 491	2 223
Career MPG	160	35 5	3 9	39 4	3063 4	19 146	706	8 9349	79 832
Career FG%	160	48	19	67	67 04	4190	0049	06261	004
Career 3PT%	160	46	00	46	36 27	2267	0115	14556	021
Career FT%	160	1 00	00	1 00	112 50	7031	0112	14228	020
Career RPG	160	10 4	0	10 4	516 5	3 228	155	1 9598	3 841
Career APG	160	7 4	0	7 4	251 3	1 571	123	1 5613	2 438
Career SPG	160	2 1	0	2 1	97 1	607	033	4147	172
Career BPG	160	2 0	0	2 0	54 9	343	031	3895	152
Career TPG	160	3 8	0	3 8	184 7	1 154	056	7120	507
Career PPG	160	19 9	1 0	20 9	1138 8	7 118	373	4 7211	22 288
NCAA MIN	159	20	16	36	4559	28 67	34	4 245	18 020
NCAA FG%	160	29	39	67	77 04	4815	0048	06022	004
NCAA 3PT%	160	46	00	46	41 41	2588	0114	14425	021
NCAA FT%	160	47	43	89	112 71	7044	0067	08512	007
NCAA PPG	160	18 5	5 4	23 9	2200 0	13 750	275	3 4808	12 116
NCAA RPG	160	8 7	2 1	10 8	929 7	5 810	161	2 0348	4 140
NCAA APG	160	7 6	4	8 0	375 0	2 343	127	1 6065	2 581
NCAA BPG	160	4 4	0	4 4	151 8	949	069	8730	762
NCAA SPG	160	2 6	2	2 8	194 1	1 213	047	5954	354
NCAA TPG	159	4	1	5	375	2 36	06	698	488
Valid N (listwise)	159								

The results showed that there were significant t stats for a number of statistical categories, though not all. T-stats showing confidence levels of 95% or above were found for the following categories: NCAA FG%, 3PT FG%, TPG, College Class, and SPG. This allows us to continue to strongly consider these categories as grounds/variables that can be considered for future related studies. After conducting a multiple variable regression analysis, the combination of NCAA PPG, TPG, SPG, 3PT FG%, College Class and FG% were shown to have the highest R-squared at .356. An R-squared of this amount could be considered significant enough to explain positive career PPG in the NBA.

A regression analysis was conducted that focused on identifying the variables that were best predictors for each position. The results yielded some interesting results as it was found that different variables influenced NBA success for players at different positions. Interesting variables that were not mentioned in the overall regression analysis of all players at all positions were present in the individualized analysis. This variation in categories/variables of importance could be accounted for by the fact that specific attributes hold greater significance at specific positions. It could be argued that, each set of players at position should be evaluated in this manner for the most accurate results.

Another of significant interest is the negative relationship that occurred between NBA performance and college class. It was shown that earlier a player left school the more productive an NBA career he was likely to have. This actually held

true to the common trend of NBA personnel to draft underclassmen at higher positions in the annual NBA draft.

With the results being such, we may have to seriously consider other statistical categories that could serve as barometers of success. Due to the lack of academic theory that exist on this topic, we must assume that the commonly held perception of significantly positive collegiate statistics may not hold as much significance as one may assume as they serve as moderate indicators of NBA productivity.

Top Combination Of Factors Contributing To NBA PPG

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.355[a]	.126	120	4.4386
2	.440[b]	193	.183	4.2774
3	.473[c]	.223	.208	4.2107
4	.538[d]	.290	.271	4.0402
5	.577[e]	.333	.311	3.9273
6	.597[f]	.356	.331	3.8718

a. Predictors: (Constant), NCAA PPG

b. Predictors: (Constant), NCAA PPG, COLLEGE CLASS

c. Predictors: (Constant), NCAA PPG, COLLEGE CLASS, NCAA SPG

d. Predictors: (Constant), NCAA PPG, COLLEGE CLASS, NCAA SPG, NCAA FG%

e. Predictors: (Constant), NCAA PPG, COLLEGE CLASS, NCAA SPG, NCAA FG%, NCAA 3PT%

f. Predictors: (Constant), NCAA PPG, COLLEGE CLASS, NCAA SPG, NCAA FG%, NCAA 3PT%, NCAA TPG

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.416	1.450		.287	.775
	NCAA PPG	.488	.103	.355	4.755	.000
2	(Constant)	4.947	1.877		2.635	.009
	NCAA PPG	.451	.099	.328	4.536	.000
	COLLEGE CLASS	-1.233	.341	-.261	-3.614	.000
3	(Constant)	4.586	1.854		2.474	.014
	NCAA PPG	.349	106	.254	3.281	.001
	COLLEGE CLASS	-1.249	.336	-.265	-3.719	.000
	NCAA SPG	1.499	.613	.188	2.445	.016
4	(Constant)	-6.476	3.419		-1.894	.060
	NCAA PPG	.297	103	.216	2.888	.004
	COLLEGE CLASS	-1.383	.324	-.293	-4.265	.000
	NCAA SPG	2.576	.653	.323	3.942	.000
	NCAA FG%	22.627	5.971	.288	3.789	.000
5	(Constant)	-11.521	3.687		-3.125	.002
	NCAA PPG	.187	106	136	1.763	.080
	COLLEGE CLASS	-1.591	.322	-.337	-4.942	.000
	NCAA SPG	2.216	.645	.278	3.434	.001
	NCAA FG%	33.560	6.758	.426	4.966	.000
	NCAA 3PT%	9.301	2.944	.284	3.159	.002
6	(Constant)	-8.088	3.922		-2.062	.041
	NCAA PPG	.270	.110	197	2.446	.016
	COLLEGE CLASS	-1.678	.320	-.356	-5.249	.000
	NCAA SPG	2.743	.675	.344	4.062	.000
	NCAA FG%	30.096	6.827	.382	4.409	.000
	NCAA 3PT%	8.514	2.922	.260	2.914	.004
	NCAA TPG	-1.296	.557	.191	-2.327	.021

a. Dependent Variable: Career PPG

Regression With All Factors Considered

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.630[a]	.396	.338	3.8513

a. Predictors: (Constant), WEIGHT, COLLEGE CLASS, NCAA PPG, NBA POS, NCAA TPG, NCAA BPG, NCAA FG%, NCAA FT%, NCAA SPG, NCAA RPG, NCAA 3PT%, HEIGHT, NCAA MIN, NCAA APG

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	8.263	15.882		.520	.604
	NCAA MIN	-.200	.161	.180	-1.243	.216
	NCAA FG%	25.002	7.738	.318	3.231	.002
	NCAA 3PT%	8.801	3.378	.269	2.605	.010
	NCAA FT%	3.937	5.157	.071	.763	446
	NCAA PPG	.317	.180	.230	1.762	.080
	NCAA RPG	.463	.268	.198	1.729	.086
	NCAA APG	.808	.466	.275	1.736	.085
	NCAA BPG	.281	.538	.052	.523	.602
	NCAA SPG	2.374	.809	.298	2.934	.004
	NCAA TPG	-1.892	.815	-.279	-2.323	.022
	NBA POS	.258	165	117	1.569	.119
	COLLEGE CLASS	-1.603	.333	-.340	-4.815	.000
	HEIGHT	-.267	198	193	-1.352	.178
	WEIGHT	2.365E-02	.021	.134	1.134	.259

a. Dependent Variable: Career PPG

Regression With Only PG

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	NCAA FG%		Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: Career PPG

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.525[a]	.276	.251	3.9428

a. Predictors: (Constant), NCAA FG%

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-21.987	8.773		-2.506	.018
	NCAA FG%	67.191	20.233	.525	3.321	.002

a. Dependent Variable: Career PPG

Regression With Only SF

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	COLLEGE CLASS		Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: Career PPG

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.495[a]	.245	.195	4.3200

a. Predictors: (Constant), COLLEGE CLASS

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	15.275	3.582		4.264	.001
	COLLEGE CLASS	-2.254	1.022	-.495	-2.206	.043

a. Dependent Variable: Career PPG

Regression With Only PF

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	NCAA PPG		Forward (Criterion: Probability-of-F-to-enter <= .050)
2	COLLEGE CLASS		Forward (Criterion: Probability-of-F-to-enter <= .050)
3	NCAA FG%		Forward (Criterion: Probability-of-F-to-enter <= .050)
4	NCAA SPG		Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: Career PPG

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.430[a]	185	162	4.4825
2	.564[b]	.318	.278	4.1605
3	.695[c]	.483	.436	3.6780
4	.742[d]	.551	.495	3.4807

a. Predictors: (Constant), NCAA PPG

b. Predictors: (Constant), NCAA PPG, COLLEGE CLASS

c. Predictors: (Constant), NCAA PPG, COLLEGE CLASS, NCAA FG%

d. Predictors: (Constant), NCAA PPG, COLLEGE CLASS, NCAA FG%, NCAA SPG

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.588	2.982		-.533	.598
	NCAA PPG	.627	.222	.430	2.818	.008
2	(Constant)	3.136	3.321		.944	.352
	NCAA PPG	.627	.206	.430	3.037	.005
	COLLEGE CLASS	-1.534	.596	-.365	-2.574	.015
3	(Constant)	-13.470	5.905		-2.281	.029
	NCAA PPG	.597	.183	409	3.266	.003
	COLLEGE CLASS	-1.735	.530	412	-3.271	.003
	NCAA FG%	33.861	10.448	409	3.241	.003
4	(Constant)	-13.431	5.589		-2.403	.022
	NCAA PPG	.432	.188	.296	2.291	.029
	COLLEGE CLASS	-1.819	.503	-.432	-3.613	.001
	NCAA FG%	32.163	9.917	.389	3.243	.003
	NCAA SPG	3.870	1.758	.286	2.202	.035

a. Dependent Variable: Career PPG

Regression With Only SF/SG

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	WEIGHT		Forward (Criterion: Probability-of-F-to-enter <= .050)
2	NCAA FG%		Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: Career PPG

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.658[a]	.433	.404	3.4090
2	.816[b]	.666	.630	2.6852

a. Predictors: (Constant), WEIGHT

b. Predictors: (Constant), WEIGHT, NCAA FG%

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-49.310	14.999		-3.288	.004
	WEIGHT	.266	.068	.658	3.905	.001
2	(Constant)	-82.297	14.893		-5.526	.000
	WEIGHT	.256	.054	.631	4.749	.000
	NCAA FG%	75.047	20.629	.483	3.638	.002

a. Dependent Variable: Career PPG

Regression With Only SF/PF

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	HEIGHT		Forward (Criterion: Probability-of-F-to-enter <= .050)
2	NCAA SPG		Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: Career PPG

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	1.000[a]	1.000	.999	.1698
2	1.000[b]	1.000	1.000	.

a. Predictors: (Constant), HEIGHT

b. Predictors: (Constant), HEIGHT, NCAA SPG

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	167.065	2.633		63.445	.010
	HEIGHT	-1.965	.033	-1.000	-59.005	.011
2	(Constant)	160.842	.000		.	
	HEIGHT	-1.892	.000	.963		
	NCAA SPG	.283	.000	.041	.	.

a. Dependent Variable: Career PPG

Regression With Only PG/SG

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	NCAA PPG		Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: Career PPG

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.719[a]	.517	.448	4.7113

a. Predictors: (Constant), NCAA PPG

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-24.013	12.575		-1.910	.098
	NCAA PPG	2.164	.791	.719	2.736	.029

a. Dependent Variable: Career PPG

Conclusion

In terms of acquisition of players, this further complicates the process for organization looking to make the best possible investment in a player. Through research conducted in this study, it has further reinforced that there is no one method by which to place complete evaluation of a player/investment on. An organization that is looking to carry out the evaluation process thoroughly should use multiple methods of analyzing prospects in order to make the best possible decision. This thesis does, though, reinforce the need to legitimately pay attention to the significance of methods like the one utilized in the thesis that derive information from statistical analysis. Organizations may also want to invest in further statistical studies that may develop formulas that can be used to target and identify players who fit the team scheme or environment. In order to maintain a competitive advantage, an organization must employ tactics that allow them to see information in ways that competitors cannot. Current employed tactics are psychological tests and IQ tests that measure the mental and emotional stability of players. These methods have become common place though as all the competitors in the NBA employ these tests and gather similar information in the process. Because of the nature of statistical analysis, one can employ numerous methods that can help gather insight and expert knowledge to eliminate the risk level of any potential investment/draft pick. My hope is that a thesis like such works to initiate future research that could lead to the improvement of the player evaluation process.

References

Onwuegbuzie, Anthony. (1998) Factors Associated with Success Among NBA Teams, Valdosta State University, Valdosta, Georgia.

Berri, David J. "Who is 'most Valuable'? Measuring the Player's Production of Wins in the National Basketball Association." Managerial and Decision Economics 20.8 (1999): 411-27. <http://links.jstor.org/sici?sici=0143-6570%28199912%2920%3A8%3C411%3AWI%27VMT%3E2.0.CO%3B2-X>

Fizel, John L., and Michael P D'Itri. "Managerial Efficiency, Managerial Succession and Organizational Performance." Managerial and Decision Economics 18.4 (1997): 295-308. <http://links.jstor.org/sici?sici=0143-6570%28199706%2918%3A4%3C295%3AMEMSAO%3E2.0.CO%3B2-%23>

Gietscher, Steve. "Playing the Numbers Game: Now More than Ever, the Sports World is Looking to Statistics for Performance-Enhancing Insight, Fueling the Quest to Devise Perfect Predictors of Success." The Sporting News 2005: 1-16. Google Scholar. Gale InfoTrac.

Groothius, Peter A., James R. Hill and Timothy Perri. "Early Entry in the NBA Draft: The Influence of Unraveling, Human Capital, and Option Value." Appalachian State University, 2005. Google Scholar.

Rosenbaum, Dan T. It Doesn't Pay to be Young in the NBA. Masters University Of North Carolina at Greensboro, 2003.

Shoten, Dan. Athletic Performance through the Determination of Brain Type., 2004. Google Scholar. westhamptonbeach.k12.ny.us.