DIMENSIONALITY REDUCTION: PRINCIPAL COMPONENT ANALYSIS OF WORKLOAD OF AN NBA GAME

by

Brian Audelio Serrano

A Dissertation Presented in Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

Liberty University, Lynchburg, VA

2023

APPROVED BY:

Jon Davis, Ph.D., CSCS, Committee Chair

David Barton, Ph.D., Committee Member

Justin Kilian, Ph.D., CSCS, Committee Member

ABSTRACT

The worlds of sports science, data analytics, and sports performance are at a crossroads. Each field is accustomed to functioning in relative silos; however, this approach is quickly becoming outdated. Sporting organizations are adopting high-performance models that integrate all departments and their subsequent data pipelines. This has led to a wealth of information not available before to a team's performance health staff in guiding optimal athletic performance. Within the sport of basketball, these silos can be split further into video coaching, basketball analytics, sports science, player development, and performance health. Specifically, basketball sports science can be categorized into internal and external workloads. Internal workload measures variables like heart rate (max, average), heart rate variability (HRV), and sleep scores. External workload can be measured using force plates, isometric strength testing, and local positioning systems (LPS). LPS systems provide a wide variety of metrics which can exceed 100 variables per player and per game. This can make data aggregation difficult and overwhelming, causing a hinderance in coaching decisions. Thus, this study used a principal component analysis (PCA) to reduce the dimensions of an LPS system on external workload. All 82 games from the 2022-2023 regular NBA season were used for this project. A PCA with corresponding measures of sphericity and collinearity was conducted using Statistical Package for the Social Sciences (SPSS). The results of this study will help clinicians and practitioners be able to understand which post-game variables are important to make decisions in guiding recovery and promoting optimal court performance.

Keywords: External Workload, Sports Science, Strength and Conditioning, Workload Management, Multivariate Statistics

Dedication

This Dissertation is dedicated to my son Carter Audelio Serrano and wife Jackie Serrano for being the light on this journey.

Acknowledgments

I would like to acknowledge Liberty University for giving me the opportunity to complete this degree in a format that allowed me to work full time simultaneously. Also, to Drs. Jon Davis, David Barton, and Justin Kilian for their guidance and mentorship through this rigorous process.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abalakov Jump (AJ)

One-Way Analysis of Variance (ANOVA)

Autonomic Nervous System (ANS)

British Association of Sport and Exercise Sciences (BASES)

Collective Bargaining Agreement (CBA)

Countermovement Jump (CMJ)

Certified Performance and Sport Scientist (CPSS)

Certified Strength and Conditioning Specialist (CSCS)

Comma Separated Value (CSV)

Drop Vertical Jump Test (DVJ)

Electromyography (EMG)

Electronic Performance and Tracking Systems (EPTS)

Erector Spinae (ES)

Eastern Standard Time (EST)

Flight Time to Contraction Time (FT:CT)

Functional Principal Component Analysis (fPCA)

Global Positioning Unit (GPS)

Ground Reaction Forces (GRF)

Heart Rate (HR)

Hear Rate Variability (HRV)

Isometric Mid-Thigh Pull (IMTP)

Inertial Measurement Unit (IMU)

Kaiser Meyer Olkin (KMO)

Local Positioning System (LPS)

Linear Position Transducer (LPT)

Major League Baseball (MLB)

Medial Gastrocnemius (MG)

National Basketball Association (NBA)

National Basketball Players Association (NBPA)

National Collegiate Athletic Association (NCAA)

Not Positive Definite (NPD)

National Strength and Conditioning Association (NSCA)

Principal Component Analysis (PCA)

Pearson Product-Moment Correlation (PCC)

Receiver Operated Curve (ROC)

Runs Batted In (RBI)

Reactive Strength Index (RSI)

Reactive Strength Index Modified (RSI-mod)

Single Leg (SL)

Singular Value Decomposition (SVD)

Squat Jump (SJ)

Subjective Rate of Perceived Exertion (sRPE)

Second Spectrum (SS)

Statistical Package for the Social Sciences (SPSS)

Tibialis Anterior (TA)

Training Impulse (TRIMP)

Vastus Medialis (VMO)

Vertical Squat Jump (VSJ)

CHAPTER ONE: INTRODUCTION

Overview

This chapter will introduce the concepts of external load and workload management as they have become topics of discussion within professional sports, namely basketball. Within external load management, there are various ways to interpret data which may become more complex as the amount of information increases in volume. As the methods to measure external workload increase, so do the ways to analyze them. The principal component analysis (PCA) will also be explored as a structured and efficient way to better understand large volumes of data. The PCA is a multi-variate statistical technique used for dimensionality reduction and will be applied to a regular season of the NBA (National Basketball Association). The drive for increasing analytics, insights, and innovation makes this major league in the United States a prime candidate for potential information overload, which may hinder the decision-making process. Thus, providing an overview as to why the NBA would benefit from dimensionality reduction will be highlighted.

Background

The world of data analytics has recently gained popularity and has been made possible through the increasing number of technical resources made available to teams and organizations. Data analytics has become popular because it provides an avenue for collected data to leverage sport performance and increase the odds of winning in competition (Watanabe et al., 2021). This field has become known as 'sports analytics' and more recently 'sports science' due to the evolving technical demands and specific skill sets needed for technical proficiency in the field. Sports Science is a unique field because it combines the knowledge of strength and conditioning such as coaching athletes and cueing movement, with technical skills seen in traditional data analytics such as statistical interpretation and data visualization. Sports science stands at the crossroads of sports medicine, sports performance, and data science (Bishop, 2008). A sports scientist uses data to leverage human performance on and off the competition floor. Sports science can be used at any level of sport but is most commonly seen at higher levels of competition such as the National Collegiate Athletic Association (NCAA) Division 1 and professional sport due to technological demands, technical requirements, and financial resources needed. Popular examples of sports science can be seen in ‘Moneyball’ where data were used to help the Oakland Athletics (Major League Baseball Team) make smarter signing decisions while still preserving a lower salary cap by using sabermetrics (advanced baseball statistics) (Mizels et al., 2022). The book highlighted how general manager Billy Beane began looking at traditionally undervalued metrics such as ‘on-base percentage’ against traditional metrics such as RBI (runs batted in) to build a successful team that was ‘cheap’ in comparison to other MLB clubs using these novel metrics as player indicators of success. More recently, The University of Miami Men’s Basketball team used sport science to leverage data and gain a spot in the Division 1 ‘March Madness’ tournament. The Miami men’s basketball program used Kinexon (LPS system) during training sessions and games to obtain information on overall player external workload variables. These variables included mechanical load which provides an estimate of how much external work a player has performed on the basketball court and may be used as an indirect measure of fatigue. By keeping account of player readiness and fatigue, the team was able to shift their game tactics around players and adapt these tactics to earn a spot in the final four (semi-final) of the biggest tournament in Division 1 basketball. Specifically, Dr. Tommy Otley (Physical Therapist and Sports Scientist) used the information provided by Kinexon to change their tactics. For example, prior to Kinexon the team was being outmatched in the third quarter in terms of game pace and shooting performance. Dr. Otley communicated these insights to the coaching staff who implemented necessary changes in the March Madness tournament in the first round, Sweet 16, and Elite 8 stages to win each game respectively. These insights were centered around increasing the pace of their game tactics during the last third of the game such as switching to a full-court press. The ability for the team to be rested enough has been led by changes in training and modulating volume, intensity, and frequency informed by the Kinexon system. This resulted in objective, positive changes such as increasing shooting performance to 59.2% from the field, 87.5% from the free-throw line, and making 13 of 14 shots within the last 4 minutes of each game. The plethora of data can also be seen on sporting networks like ESPN who show win-loss ratios for teams, individual player efficiency rating, and how each player affects a team's performance through their plus-minus ratings. If the general public has access to this much information, it would seem logical that individual teams would have access to at least the same amount of data points. More importantly though, each team strives to be at the forefront of technology and its insights. This makes the field of sports analytics and sports science interesting due to its ability to provide additional leverage for teams in their pursuit to win championships. The NBA is the highest level of professional basketball played in the United States and the world. Its regular season begins in late October and concludes in mid-April consisting of 82 games (41 home and 41 away) (Gonzalez et al., 2013). To provide context into player salary, rookies drafted into the league are signed to a 4-year contract worth $925,258 yearly (minimum). The NBA draft consists of 2 rounds (60 picks) with the top pick being paid $9,212,600 in the first year and decreases by 10% for every pick thereafter for the first round only. Thus, first round draft picks obtain 2 years of a guaranteed salary. With more experience, such as 10+ year veterans have yearly contracts worth at least $2.9 million. The average player for the 22-23 season made $9.2 million while max contract players are those who are paid the max amount percentage of a team's salary cap. To put this numerically, these players can earn from 25-35% of a team's total salary cap which equates to +$100 million (Lyons et al., 2015). The National Basketball Players Association (NBPA) negotiates these salary standards for players through a collective bargaining agreement (CBA). These monetary values show just how much each player is worth, making them huge financial assets to each team and their stakeholders. This is partially where the competition for data analytics comes into play, in giving teams every possible advantage to win championships and increase organizational value/revenue.

Data Analytics has a place in professional sports because as mentioned earlier, the resources available at this level are nearly limitless. This is paired with the large financial investment made into each player ranging from personal chefs, massage therapists, and skill coaches in addition to the full-time support staff. The combination of limitless resources and large financial assets lead to a plethora of sport technologies which can include internal and external workload measures. Internal workload records the physiological response to exercise and training through tools such as heart rate variability and changes in jumping performance. External workload records how much work is being performed on the field of play or in the weight room through Local Positioning Systems (LPS) or Global Positioning Systems (GPS) systems. As sport technology evolves at a rapid pace, the amount of variables captured do as well. This presents a challenge to the sport scientist who may need to adapt to the changing environment from only a handful of variables to tens, and potentially hundreds of variables across different devices/programs. Kinexon is one such example where each report can yield 134 total variables for each player. One common question brought up is how to obtain meaningful and actionable data from all of this information. The field of statistics can provide one such avenue through techniques such as machine learning and multivariate statistics.

The principal component analysis (PCA) is a statistical technique used to reduce the amount of variables within a large data set (Abdi & Williams, 2010). In other words, this technique sets out to reduce data sets with many dimensions (variables) into one with fewer dimensions to make interpretation easier (Shlens, 2014). This concept of dimensionality reduction can be divided into feature selection and feature extraction (Motoda et al., 2009.). Feature selection looks for a small number of 'features' to represent the total data more accurately. It seeks to improve model performance through eliminating irrelevant or redundant variables. Feature extraction derives (extracts) information from the original dataset into a smaller amount of relevant variables. Both feature selection and extraction allow models to run faster, be more applicable, and become simpler to understand. The primary difference is that while the original data is maintained in selection, it is transformed and reduced to new variables in extraction methods. Examples of feature selection include regression and classification models. Feature extraction (or projection) strives to reduce the amount of variables in the total data set and can be linear or non-linear. The PCA is an example of a linear technique and will be used for this research study while a technique such as manifold learning can be used for non-linear data sets. The linearity of the data was investigated by a singular value decomposition (SVD) method. The SVD method places all points (variables) of the dataset onto a symmetrical and square matrix. Since each value is semidefinite (zero or positive value), the variance and covariance can be defined on a cartesian coordinate system to investigate linearity. If the assumption of linearity is violated, then a non-linear technique such as manifold learning may be better suited for analysis. When using a PCA technique, the total data is placed on a straight line and a covariance or correlation matrix (dependent on analysis type) is created and resulting eigenvectors are calculated to measure the strength of each principal component (Jollife & Cadima, 2016). In general, the larger the eigenvector, the more variance can be explained within the model and attributed to certain variables. However, the PCA comes with trade-offs such as diminished accuracy and model overfitting since the amount of data is reduced in relation to the original dataset. To provide greater insight; the following are the main steps when performing a PCA (Kherif et al., 2020). The first step is standardizing every variable so each holds equal weight in the final model (to reduce sensitivity bias). Step 2 is to create a correlation (or covariance) matrix to further understand relationships between each variable. From the matrix (correlation or covariance), eigenvalues and eigenvectors are then computed to determine the principal components. Eigenvalues help explain the variance between each principal component such that higher eigenvalues yield more prediction power in a principal component. The final principal components are then shown in a scree plot and analyzed to look at which amount of components explain the most variance. The sports scientist can then take these newfound variables as being the most pertinent from the original data set and begin making decisions using said variables as guides. Laffaye & Bardy (2007) looked at differences in one-leg running jumps and how athletes from different sports perform this movement in twenty-five male athletes of regional level ability. A PCA was conducted on 200 jumps (data points) on force plates extracted from force-time curves on force plates. They found that two principal components explained 78.3% of the variance in jump height using an Eigenvalue threshold greater than 1.0. The two primary components included temporal and force variables. The authors add to the discussion how different athletes have different jumping strategies to create height (Laffaye et al., 2007) . Volleyball players rely more on temporal factors such as decreased ground contact times and decreased eccentric contact times (force plate metric: eccentric duration in ms, total time in ms.) while basketball players use a more force-based approach. Kollias et al. (2001) looked at individual differences in vertical squat jump performance through a principal component analysis in male elite level athletes. Specifically, they set out to explore the relationship between lower extremity muscular strength and vertical jump height since previous correlations have been weak to moderate. Historically, Aragon-Vargas & Gross, (1997) used multiple regression analyses to explore these relationships and found different levels of variation depending on type of analysis used (whole-body or segmental). The method by Kollias et al. (2001) has been utilized in other studies (Laffaye & Bardy, 2007; Kleinbum, Kupper, & Muller, 1998**).** It used SPSS (version 8.0) to perform a factor analysis and apply a singular value decomposition to create principal components with Eigenvalues greater than 1.0. The authors found two principal components explained 74.12% of total variance of six force-time variables of which impulse duration and time to peak force were predominant in the first principal component (38.65% of variance). Panoutsakopoulos et al. (2014) also used a PCA to look at the vertical squat performance in female athletes throughout multiple sports. Specifically, they followed the research from Kollias et al. (2001) that athletes use force-time characteristics to different extents. This study analyzed squat jumps on force plates performed by 173 female athletes (Panoutsakopoulos et al., 2014). Data were analyzed using the raw force-time curves into force and temporal factors. A PCA with varimax rotation and Kaiser normalization was used to create principal components and verified by Cronbach's alpha for reliability. The two principal components explained 69.1% of total variance (40.2% eigenvalue 2.41, 28.9% eigenvalue 1.73 respectively). The findings show that track and field athletes have higher rates of impulse and subsequent force production while volleyball players use a more time-based approach. These studies show how the PCA can be used to find more pertinent variables in large datasets.

While traditional methods of statistics such as t-tests for continuous variables, chi-square tests for categorical variables, and correlations are common teachings in undergraduate statistics classes, more advanced techniques are not present in curriculums. To learn more advanced concepts like regressions and linear algebra, one must usually pursue post-graduate education specifically within the field of statistics or other related fields. Thus, it becomes very difficult for a clinician with a health professions-based education or a performance practitioner with a strength and conditioning education to receive this additional statistical training. This creates a large gap in how performance related data from force plates, strength assessments, and LPS information are connected with health and human performance where a sports scientist could add value to an organization.

Historically, data analytics within sports is a very young field and is not clearly defined. Gleason et al. (2023) investigated the terms of 'sport science' and 'sport scientist' and their current use in sport. The British Association of Sport and Exercise Sciences (BASES) defined sport science as "the application of sport and exercise principles within high performance sport, where the application of science is concerned with maximizing the performance of an athlete or team". Bishop (2008) added that sports science is an interdisciplinary field concerned with the understanding and enhancement of human sporting performance. Lastly, French & Torres-Ronda (2021) add how sport science is "the study of sport through the application of scientific methods to the fields of human performance, athletic endeavor, and sporting competition". For example, while the term 'sports science' may be well defined in other sports, most NBA organizations do not yet have a practitioner in this role. Those that do, still do not have a standardized title for this position unlike the field of athletic training (Athletic Trainer) or strength and conditioning (Strength and Conditioning Coach). For example, the term ranges from ‘Performance Scientist’, ‘Performance Engineer’, to hybrid roles where strength and conditioning coaches also have duties to gain insights from data. The infancy of the field also foreshadows the amount of progress to be made. Unlike sports such as Aussie Rules Football and European soccer who have entire departments dedicated to sports science, the NBA may only have one person dedicated to this role. Furthermore, the siloed information approach between departments can make obtaining actionable information difficult. For example, basketball analytics looks at information pertaining to player scoring, matchups, and overall efficiency. Video coaching looks at how individual players and opposing players move on the court against each other in response to plays and possessions. Sports Performance (or Strength and Conditioning) looks at how physical development in the weight room transitions to performance on the court. Sports Science seeks to leverage data captured on the court, in the weight room, and athletic training room to increase on-court performance.

The metrics looked at by sports scientists are vast but can be categorized into external and internal workloads. Internal workload metrics measure how the body physiologically responds to stress and are commonly measured by tools like heart rate variability (HRV), sleep scores with Oura rings or recovery scores with WHOOP bands. HRV measures the variation in duration between heartbeats (R-R intervals) and is used as an indirect measure of cardiac regulation by the autonomic nervous system (ANS) (Morgan et al., 2017). Hedelin et al. (2001) investigated how HRV affected sport performance in elite cross-country and canoe athletes. They found that athletes with increased HRV also had increases in V02max, extremity (leg and arm) peak torque, and time to peak torque. Heart Rate (HR) can also be indirectly measured using a chest strap or wrist strap and used to calculate TRIMP (Training Impulse) or Zones of Cardiorespiratory activity. The TRIMP score measures training impulse (internal workload) by multiplying length of training session and average HR which is used by the company FirstBeat (Heishman et al., 2018). Polar (Kempele, Finland) (wearable tech company) also uses maximum HR percentages to split training into five heart rate zones ranging from Zone 1 (Very Light) at 50-60% of HRmax to Zone 5 (Maximum) at 90-100%. Each intensity and zone corresponds to a different training effect and physiological quality (Sylta et al., 2014). External workload seeks to measure how much stress occurs during training and games. Common tools used in basketball include force plates, isometric strength testing, and LPS systems. Force Plates use countermovement jumps (CMJ) and their variations (Squat Jump, Abalakov, Hop Tests) to look at concentric and eccentric variables which may be correlated to fatigue and player readiness (Lombard et al., 2020). Two variables used in the CMJ and squat jump (SJ) include Flight Time to Contraction Time (FT:CT) and Reactive Strength Index-modified (RSI-mod). FT:CT evaluates the athletes jumping strategy and RSI-mod which is derived from dividing time to takeoff by jump height measures an athlete's explosiveness (Heishman et al., 2020). Any decreases in such metrics may signal a decrease in readiness status. Isometric testing of the major muscles of the lower extremity may be used to look at peak force or limb symmetries in the days leading up to a game to also establish player readiness. Merrigan et al. (2021) looked at the intrasession reliability of the isometric mid-thigh pull (IMTP) and relationship between CMJ and IMTP. They found reliability was met for variables such as Force at 150 ms, Force at 200 ms, Net Peak Vertical Force, and Peak Vertical Force; changes in these metrics may be used to assess player readiness. Lastly, LPS are used to measure total external workloads and how to best recover from training loads or game loads going into the next competition. Two popular local positioning systems used in the field and literature are Catapult (Melbourne, Australia) and Kinexon (Chicago, USA). Catapult uses Player Load (proprietary metric), which is the sum of accelerations through all planes of the tri-axial accelerometer during movement and calculates an integer to make comparison simpler (200 vs 300). Kinexon uses a similar principle in Mechanical Load (proprietary metric) which captures all acceleration and deceleration loads to produce a measure of external workload.

Although a plethora of technologies are available for measurements to measure internal stress and its response, the collective bargaining agreement (CBA) is rightfully strict when it comes to what players can wear in game and training environments. The NBPA and CBA strive to protect player health and wellness since this information could theoretically be used against the player. This has occurred in settings such that these metrics are used to quantify player effort during training or if they are working hard enough which has left a negative stigma within players and their perception of sports science. For example, teams may have any of the aforementioned technologies and testing procedures available in their practice facility, yet players are not mandated to participate, it is strictly voluntary. Games are even more strict in that players are not allowed to wear any sport technology. This brings to light the question of how teams can quantify work performed in games. Teams are able to overcome this quantification issue through Second Spectrum (SS). Second Spectrum is the official optical tracker and game analytics provider of the NBA as of the 2022-2023 season. Second Spectrum collects all game data via optical tracking and then coordinates with Kinexon to translate this data into the native environment. Thus, teams are still able to track players in games through a partnership between Second Spectrum and Kinexon which is delivered the morning after a game. The issue remains however of how to pick out the most important variables from the oversaturation of information.

Problem Statement

The challenge then becomes how to interpret this data and transform it into actionable information. The native Kinexon environment can be categorized into portable (IMU) and permanent (LPS) units (Rana & Mittal, 2021). The portable unit uses an IMU (Inertial Measurement Unit) to gather performance metrics based on a central receiver. The permanent unit places receivers throughout a practice facility and captures data through a wearable sensor and corresponding LPS system. At the completion of a training session, a customized report can be generated at the team or individual player level. Each report of the travel version consists of about 35 metrics while the permanent system can output 134 metrics. It is then the duty of the sport scientist and performance staff to interpret this large amount of information in a timely and efficient manner for the coaching and front office staff. This is where principles of dimensionality reduction may be applied such as the principal component analysis (PCA) to the produced large datasets. Historically, the analysis process of data has been left to subjective and lower-power statistical methods. For example, Alemdaroglu (2012) investigated the relationship between isometric knee strength, anaerobic performance, and vertical jump performance in 12 professional basketball players. The author used a Pearson Product Moment Correlation (PCC) analysis on the aforementioned variables and found no measure of strength was significantly related to any of the field tests. Brumitt et al. (2021) took the approach of injury prevalence with basketball plus volleyball and Reactive Strength Index (RSI) using the drop vertical jump (DVJ) test. The authors found volleyball players with an RSI of less than 0.9125 m/s were 4 times (relative risk 4.2, 95% CI: 1.0, 17.7; p-value: 0.024) more likely to be injured. The RSI was calculated by dividing jump height and ground contact time in the DVJ. Receiver operated characteristic (ROC) curves were used to classify athletes into groups of which cutoff scores were used to calculate the relative risk of injury. The primary concern with correlational statistics is that it cannot predict causation. Similarly, using ROC for diagnostic accuracy may lead to calibration errors of the C statistic and subsequent errors in sensitivity and specificity. Alternatively, clinicians and practitioners have simply looked at either datum or players in search for outliers. For example, if an athlete has an average workload of 800 AU (arbitrary units) and one game results in a spike, this may invoke a premature response from the performance health staff to unnecessarily rest a player (Wang et al., 2020). Beyond this, programs like Microsoft Excel have been used to find summary statistics of player workload and performance metrics. Several issues arise from measuring external stress from subjective means like overestimation or underestimations. While summary statistics are a great place to begin data investigation, they do not provide answers to higher level questions like prediction or relationships between different variables.

Purpose Statement

The purpose of this study was to reduce the dimensionality of Kinexon data through the 2022-2023 NBA season. Since each Kinexon report can produce 134 variables per player it may become very difficult to glean actionable information from each player report. Currently, the research is scarce looking into how sport technology information can be interpreted. Thus, by performing a PCA on these large datasets it may become easier to gain insight and make evidence-informed performance decisions. The PCA allowed for data to be reduced into 3 or fewer principal components and visualized into a scree plot and correlation matrix to give the most important variables pertaining to this unique population of athletes. This dissertation was able to reduce the original data set of 134 variables into 15 metrics of most importance.

Significance of the Study

Since a PCA has not been performed in the NBA population at the time of writing this dissertation, there is a wide variability in what each team and organization deems important in their decision-making process. For example, Team A uses total accelerations and total decelerations as their primary point of analysis while Team B uses total distance and mechanical load. While the nuances of individual team needs should be appreciated, it seems there is a large gap of knowledge in this area. Stone et al. (2002) partially investigated this issue within training sessions in male Division 1 Basketball players. The authors found Max Speed, Total Decelerations, Total Jumps, and Total Mechanical Workload (proprietary measurement that summates external workload performed by a player) were the main variables of interest found through the PCA. Parmar et al. (2018) used the PCA to develop relevant performance metrics in professional rugby. The authors found 10 principal components explained 81.8% of variation. Of the 10 principal components, Possession and Speed of Play accounted for 62% of variance with focus on ‘amount of possession’ and ‘making quick ground’. Furthermore, this topic becomes extremely important as practitioners and clinicians make decisions into workload management, recovery, and training. The NBA plays an average of 3.5 games per week, making what occurs on non-game days vital to optimal performance on the court. Within game scheduling, there is the potential for back-to-back games along with being on road trips. The combination of high workloads, short recovery periods, and circadian rhythm disruptions due to travel make workload management a prime topic of conversation within the NBA. By being able to discern the most important variables of load management within Kinexon, practitioners and clinicians can better help prepare for the challenges of the season.

Research Question(s)

RQ1: Which external workload variables have 70% or more of their variance explained by the principal components (communalities extraction value 0.7 of greater)?

RQ2: Will greater than 70% of total variance be explained by 3 or fewer principal components with Eigenvalues greater than or equal to 1.0?

RQ3: If 3 or fewer components explain 70% or more of workload variance-is there a distinct combination of variables having a component matrix loading score of 0.4 or greater that describe the components?

Definitions

1. *Kinexon*- A local positioning system that measures athletic external workload (Rana & Mittal, 2021)
2. *Principal Component Analysis (PCA)*- Statistical technique used to reduce the dimensionality of large datasets (Jollife & Cadima, 2016)
3. *Workload:* The amount of work performed by an athlete, can be quantified in various ways (Benson et al., 2020)
4. *Internal Workload*: How the body responds physiologically to stress/work of a session. (Seshadri et al., 2019)
5. *External Workload*: How much stress/work in placed on the body and measures in objective units such as total distance covered or max speed achieved. (Seshadri et al., 2019)

CHAPTER TWO: LITERATURE REVIEW

Overview

This chapter will provide an overview as to how the PCA was applied in the basketball setting. First, the technical underpinnings of the PCA as a statistical technique were explored along with its origin in mathematics. The PCA was then explored in various settings to gain a better understanding of how and why it is used. The focus then shifted from a general understanding into the specific ecosystem of athletics and sport. Finally, the PCA explored the context of basketball and how it may be applied to assist with performance-based decisions. A summary was provided at the end of chapter that demonstrated the PCA as a technique along with its value from general sports performance to specific basketball decisions.

Conceptual or Theoretical Framework

The principal component analysis is a dimensionality reduction method that is used to reduce the dimensions of large data sets, it was first used by mathematician Karl Pearson in 1901 (Tharwat, 2016). This method is strengthened by being able to preserve the accuracy and integrity of the original data set. For this to occur, there are generally 5 steps necessary in the process: standardization, computing a covariance/correlation matrix, computing eigenvalues/eigenvectors, creating a feature vector, and observing the data along the principal axes (Chatfield & Collins, 1980). Standardization strives to normalize each variable so each one contributes equally to the final model. Specifically, standardizing variables reduces the risk of data bias in variable spacing. For example, variables with larger ranges would result in greater influence over those with smaller ranges. This issue is overcome through having each variable being scaled equally through conversion into a z-score ((value-mean)/standard deviation) where each variable has a mean of 0 and standard deviation of 1. Step 2 is computing a covariance/correlation matrix to help reduce redundant variables. The purpose of the correlation matrix is to explore if any relationship exists (and to what degree) between the variables since highly correlated variables could theoretically create their own principal component and take up necessary space in the final model. This step is two-fold, one being to see if any redundant information exists between highly correlated variables and to establish a baseline between all variables in the data set. Once the correlation matrix has been calculated, the direction is observed (positive, negative, or zero). For example, a positive correlation indicates a direct correlation while a negative correlation indicates an inverse correlation (Dziuban & Shirkey, 1974a). Step 3 computes the eigenvectors/values based on the covariance/correlation matrix. The eigenvalue is a concept taken from linear algebra and is a special set of scalars associated with a matrix equation. Eigenvalues are used to study different equations and continuous dynamical systems (Larsen & Warne, 2010). Eigenvalues are used to represent the line length and amount of variance from the new axes. Each eigenvalue is paired with a corresponding eigenvector. The resulting eigenvector signifies the orientation of the new axes. Thus, it is the eigenvectors that will be projected onto the new axes and referred to as principal components. These two concepts apply to the PCA because every correlation matrix will produce a related eigenvalue and eigenvector. Once the eigenvalues and eigenvectors have been computed, the principal components are identified. Principal components become new variables that condense variables of the original dataset into new ‘principal components’ through linear combinations. The purpose of principal components are to condense the large number of variables and find which of them explain the largest amount of variance (Holland, 2019). It is important to note that each principal component is unrelated to each other which aids in data interpretation. A scree plot can be used to visualize each principal component and their eigenvalue in descending order from first to last depending on the amount of original variables. The scree plot is an important part of seeing inflection points as the amount of variance explained decreases. Once all principal components have been constructed, decisions to keep or discard a principal components are made on the amount of variance explained. The total variance for each principal component is summarized in a feature vector. The feature vector is a matrix that contains all eigenvector values which will be kept in the final model. The final data set will by definition have less dimensions since principal components with small eigenvalues are discarded (Jolliffe, 1972). The final model re-plots each principal component (based on eigenvalue thresholds of >1.0) that contain the most amount of information and variables for the analyst to use as data points. Lever et al. (2017) describes the PCA as a statistical technique that simplifies the complexity in multi-dimensional data while retaining the integrity of the original information. The authors go on to describe this method as unsupervised machine learning and is similar to clustering. Briefly, the PCA can be used as a method of machine learning. Machine learning is a process where a system learns how to improve a dataset through training algorithms without direct human control. As it applies to sport, machine learning may be used to make predictions and provide insights from live data (Richter et al., 2021). Machine Learning can be supervised or unsupervised. Supervised machine learning is done by an analyst who trains the data set and relating algorithm to find specific patterns. Unsupervised machine learning learns directly from a dataset without any human input or interaction. Since a PCA seeks to group together data points with similar data, it is closely related to clustering and unsupervised machine learning (Koshkina et al., 2021) When a data set is graphed onto a 2-dimensional axis, it may seem overwhelming and unrelated. A PCA strives to capture several lines of best fit through the rotation of x-y axes to create principal components such that each component is orthogonal to each other. These principal components are meant to capture and explain the variance in the data to reduce the dimensionality of the data. The integrity of the data are kept in part due to the met assumption that each principal component is unrelated to the next and redundant variables from the original dataset were excluded. When a dataset is graphed on a 2-dimensional axis, it again may appear overwhelming and difficult to interpret. Assumptions that no variables are related must be met before the principal components are calculated (Dascalu et al., 2017). Once this assumption has been met, principal components are lines of best fit which strive to capture as much variance from the original data set. It is important to note that each principal component is orthogonal to the other and thus are not related. Another note is that although a PCA may seem like multiple linear regressions, the PCA technique differs in that it minimizes the perpendicular distance between a data point and the principal component, whereas linear regression minimizes the distance between the response variable and its predicted value. Kambhatla & Leen (1997) also point out the limitations of using a PCA; these include the inability to capture nonlinear patterns, nonorthogonal patterns may be underrepresented, and not being able to differentiate between close clusters on the x-axis. Jim Frost also describes the PCA as a technique for reducing very large data sets into smaller indices (Principal Component Analysis Guide & Example - Statistics By Jim, 2023) . Specifically, each principal component created is a set of correlated variables which are linear combinations of the original variables. Each principal component strives to capture as much information into each index with the first principal component explaining most of the variance for each data set. Frost goes on to explain the similarities between a PCA and Factor Analysis since both contain similar theoretical mechanisms. The factor analysis identifies latent variables that cause the observed values of outcome variables. This process is guided by the analyst; thus, the point is to maximize a conceptual understanding. The PCA on the other hand reduces the dimensionality of a data set with the primary goal being to retain information rather than conceptual understanding. This is where a PCA relies on its principal components and covariance/correlation matrix to find importance in its variables. One strength of using a PCA is the reduced potential for model overfitting which occurs as variables begin to increase (Suhr, n.d.). Another obstacle becomes if variables from the original data are highly correlated since this also increases a model's error and causes subsequent loss of statistical power. Using a PCA addresses most of these challenges, for example, when dealing with more features than observations, variables are reduced while the integrity of the information is still retained. Each principal component produces independent components which address the issue of multicollinearity. In a final model, the highest principal components are graphed to visualize and understand data. This is helpful in eliminating noise since principal components with small variance values are discarded. From a visualization standpoint, each principal component is rotated to capture the highest amount of variance from all existing variables and provide the reader with a new vantage point. However, each principal component can't be interpreted yet, this is where eigenvectors and eigenvalues come into play. Each principal component has a corresponding eigenvector to identify the orientation of the new axes and eigenvalue to represent the amount of variance the new rotated axis explains. Both of these values are calculated from the preceding correlation matrix. When using a statistical software, eigenvectors will be arranged from largest to smallest in alignment with the amount of variation explained. The numerical value of a 1.0 Eigenvalue is traditionally used as a threshold for deciding when to include a principal component or discard it (Johnstone, 2001). Variables can then be analyzed within each loading plot to see the highest value of correlation. These values can be visually displayed onto a loading plot containing the first few principal components. Each principal component can be interrogated and graphed based on its eigenvalue, scree plot, and loading plot. As it relates to basketball, a Kinexon report contain 134 variables for every player, per game which can make finding the most pertinent variables difficult. Thus, the next section explores how the PCA technique has been used within sports in general to find these important variables of performance. Once the PCA has been supported as a viable technique, the sport of basketball will be explored since that is the focus of this project.

Related Literature

General

This section of the chapter will include related literature where the PCA has been used in athletic performance. Charoenpanich et al. (2013) investigated which muscles are recruited most in the vertical jump through a PCA as it relates to volleyball performance. Twenty subjects (10 volleyball players, 10 non-athletic controls) were used to perform a squat jump (SJ) and vertical stop jump (VSJ). The VSJ is specific to the sport of volleyball and is the combination of a running start with a CMJ. This study built on previous knowledge which found the primary contributors in the SJ were the knee extensors while in the VSJ added the ankle plantar flexors and upper extremity (Chappell et al., 2007; Spägele et al., 1999). Each subject wore electromyography electrodes (EMG) throughout the body to obtain an indirect measure of muscle excitability while they jumped on force plates. The force plates however were only used to obtain ground reaction forces while a yard stick was used to measure height. This style of SJ is interesting because it is usually performed with the hands-on hips throughout the entire movement to reduce both countermovement and influence of the upper extremity. When performing a PCA, a correlation or covariance matrix may be calculated, this study used a covariance matrix. This study found the first principal component explained 55% of total variance and was the only one reported. In the volleyball group, the vastus medialis (VMO), tibialis anterior (TA), and erector spinae (ES) were used the most while in the VSJ it was the TA, medial gastrocnemius (MG), and ES. The PCA was found to be a more sensitive analysis than examining the area under the curve when exploring differences between the control and experimental group. This study made the novel findings regarding the role of the upper extremity (latissimus dorsi and trunk extensors) in jumping ability within the sport of volleyball which may serve useful in strength and conditioning programming. James et al. (2021) sought to reduce the dimensionality in countermovement jump (CMJ) metrics for ease of interpretation by practitioners. The CMJ is a versatile movement and test because it can be used to guide programming, measure performance, and monitor fatigue indirectly. However, the amount of data output from each CMJ can make it difficult to discern information from noise, which is why a PCA was used to reduce the dimensionality in this data set. CMJ data was collected from 3 cohorts (recreationally trained males[1], competitive MMA athletes [2], junior elite Australian rules football players [3]). Three separate devices were also used: force platform (Bertec), linear position transducer (LPT) (MuscleLab), and force plates (ForceDecks, VALD, Australia). Data was tested for sphericity using Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure. Bartlett's test of sphericity interrogates a dataset to see if the correlation matrix is significantly different from an identity matrix (Tobias & Carlson, 2010). In other words, it checks for correlations in the present variables since this is needed to proceed with a factor analysis. The KMO measure of sampling adequacy test examines the strength of a certain correlation between variables, with values closer to 1 signaling dimensionality reduction can commence (>0.6 being the common threshold used). The amount of variance and amount of relevant principal components will vary based on the original dataset. In the first dataset, average power and rate of force development were most pertinent. In the second dataset, average power and peak power were the most pertinent. And in the third dataset concentric mean power/BM and peak power/BM were most pertinent. The authors concluded these variables could be used as target variables to examine CMJ performance. Wu et al. (2019) expanded on the research by James et al (2021) by focusing on the CMJ and fatigue monitoring. The authors state the traditional PCA method can be challenged when the number of features returns an incomplete result and the dimensionality of data exceeds the number of samples. The functional PCA (fPCA) overcomes this challenge by using the force-time curve of a CMJ as a function rather than as a sole variable. Ten recreationally trained males underwent three repeated sprint protocols (low, moderate, high volume) and performed 5 CMJ's before and at specific time intervals after the training session (Wu et al., 2019). The authors used a variation of the CMJ where a standard weightlifting bar (20kg) was held on the shoulders throughout the movement. For the data analysis, it is important to note only concentric variables were examined. Two principal components explained 68.7% of total variance. It was found that time points up to 1-hour post training were related to metabolic fatigue while 3-48 hours post training were related to neuromuscular fatigue. Floria et al. (2018) looked at how training affected force-time curves using a PCA on CMJ metrics. Floria et al. (2018) describe the PCA as an orthogonal transformation technique which converts correlated variables into a smaller number of uncorrelated variables called principal components based on the work by Deluzio & Astephen (2007). This study used thirty-four female basketball players who underwent 6 weeks of complex training (resistance training + plyometric exercises). Within the CMJ, the downward and upward phases were used for analysis of all curves (force-, velocity-, displacement-, RFD-time curves). The authors followed the approach set forth by Deluzio & Astephen (2007) and created a matrix (correlation vs covariance not specified) which resulted in eigenvectors/eigenvalues. In the force-time curve, 3 principal components accounted for 83% of variance, in the velocity-time curve 2 principal components accounted for 81% of variance, and in the displacement-time curve 1 principal component accounted for 78% of variance. Since three different time curves were analyzed, the authors were able to make the following conclusions: The training group was able to produce more force at the start and end of the upward phase of the vertical jump. In the velocity-time curve, the training group produced a faster concentric takeoff velocity along with a deeper countermovement in their displacement. Welch et al. (2019) began looking at the relationship between the CMJ and football performance using a PCA. Twenty-five football athletes underwent a variety of cutting, isometric testing (IMTP), and explosive strength testing (single-leg SJ, single leg drop jump). Each participants performed the tests on force plates while wearing reflective markers for kinematic and kinetic data. The PCA found the SL drop jump with 110 degree cut performance test was best described by RSI and jump height. The SL drop landing with 100 degree cut performance outcome was best described by a 250 ms impulse. The authors discussed how performance on single leg strength tests measuring ground reaction forces (GRF) also had an intricate relationship with cutting performance as it relates to soccer. Pino-Ortega et al. (2021) expanded on using the PCA to find the most relevant metrics in soccer, basketball, and rugby using wearable technology as they relate to talent identification. The authors approached the problem through a systematic review that included data from 34 articles (17 soccer, 11 basketball, 6 rugby). The extracted variables were then clustered into five overarching metrics: technical, tactical, biomechanical, physical/physiological, and anthropometrics. In soccer, 54-81% of total variance was explained by less than 23 variables. In basketball, 62% of total variance through less than 14 variables. In rugby, 52-90% of variance was explained by 9 variables. This study makes the conclusions that in soccer, anthropometrics and a players relative age are important to success. In basketball, anthropometrics and agility tests are important in talent identification. In looking to create connections beyond force plates and individual metrics, McCormack et al. (2021) used a PCA to compare physical characteristics in youth rugby players (McCormack et al., 2021). The authors describe the PCA technique as a way to reduce data multicollinearity and condense the numerous variables into fewer principal components. Each principal component contains variables which explain a certain portion of total variance. Six hundred and fifty-four males were recruited for this study and measures of performance included anthropometrics, CMJ, IMTP, Speed/Momentum (40-m speed test), and a yo-yo test for intermittent recovery. McCormack et al. (2021) implemented the PCA technique used previously by Weaving et al. (2018) and Scantlebury et al. (2020), interestingly, missing data was imputed using a probalistic PCA. Principal components 1 and 2 explained 69.4% of total variance. PC1 showed max momentum,10m momentum, and IMTP were most important while PC2 showed 40m speed and the yo-yo test were important physical qualities in rugby. These findings showed that running speed, absolute strength, and rate of force development are important qualities for player development. Rojas-Valverde et al. (2021) began looking at player performance in soccer through electronic performance and tracking systems (EPTS) or LPS. This study sought to apply the PCA technique onto IMU devices and the external workload variables they captured in two professional soccer tournaments. This study did use a threshold of playing more than 85% of the whole match (>76.5 min) and each match period (>38.25 minutes) which resulted in 17,674 individual data points. The IMU system used in this study was the WIMUPRO (RealTrack Systems, Almeria, Spain) and included an accelerometer, gyroscope, and GPS system. Within the data analysis, the Kolmogorov-Smirnoff test confirmed the normality of the data to find six principal components. Of the six principal components, three explained 58.75-61.67% of total variance. A mixture of distance, speed, and acceleration were found to be the most pertinent variables of external workload. Within the first principal component (27.92-30.99% of variance) was made up of relative distance, explosive distance, and high speed which may allow practitioners to focus on these metrics as main performance characteristics in professional soccer. Weaving et al. (2013) used this same approach within rugby. Seventeen professional rugby players were tracked over two, 12-week preseason periods using GPS and subjective rate of perceived exertion (sRPE). The authors used a Pearson correlation matrix to inspect the factorability of the data along with the KMO method for sampling adequacy. Within this pre-season period, the first principal component explained 68% of total variance and consisted of small-sided games while the second principal component consisted of strength and speed qualities. This article gives insight into how using both internal and external measures of workload can be used to tease out important qualities to focus on during the pre-season period.

PCA Within Sport

There have only been a few studies which specifically address how the PCA technique can be applied to metrics of workload in sport. Saucier et al. (2021) used compression shorts (Strive Sense 3, Seattle, WA) to measure external workload (Saucier et al., 2021). The Strive shorts represent a novel tool in workload monitoring because it adds surface electromyography (sEMG) to accelerometry and GPS capabilities. The Strive shorts were worn by 15 Division 1 Basketball players in training and regular season games. Interestingly, players were grouped into their positions (Guards, Forwards, and Centers) for the analysis. The authors explain how principal components are independent and orthogonal to each other, corresponding to the direction of variation within the data. This creates a new dataset with variables of interest condensed into the principal components while maintaining the integrity of the original dataset. For each position and environment (training vs game), two principal components accounted for more than 70% of total variance seen. For Centers, game sessions (defined as live competition play) were explained by accelerations, decelerations, and total distance. For Forwards, game sessions were explained by the accelerations, high accelerations, and number of jumps. For Guards, game sessions were explained by accelerations, decelerations, and total distance. In regard to Muscle Load and muscle excitability, Guards and Forwards had a higher workload than Centers. When it came to muscle load, Centers and Forwards had a higher muscle load in games for all variables vs Centers. Stone et al. (2022) performed a similar study with the intent of simplifying external workload data in Men's collegiate basketball. The authors describe the PCA as a technique which identifies multiple variables with low multicollinearity that explain high level of variances. Ten Division 1 men's basketball athletes were categorized in their positional groups (Guards, Forwards, and Centers) and given IMU sensors (Kinexon, Munich, Germany) to be worn during training and games throughout the season. Sampling adequacy of the data was performed using a KMO test with value above 0.6 as a threshold for the PCA to be conducted along with Bartlett's test to confirm divergence of correlation and identity matrices. The first two principal components explained 81.30% of total variance in the basketball athlete's workload throughout the season (Eigenvalues > 1.0). In the first principal component, total decelerations, total accelerations, average speed, total mechanical load, and total jumps were the main contributors while max speed was the primary contributor in the secondary principal component. In regard to external loads, the primary metrics of difference included max speed, total decelerations, and total jumps.

Summary

The PCA technique has been established as a means to reduce the dimensionality of large data sets. This is accomplished by taking large data sets and analyzing them to meet certain standards. For example, commonly used tests include Bartlett's test of Sphericity and the KMO measure of sampling adequacy are performed on the data to ensure variables are correlated. Bartlett's test identifies whether a matrix of correlations is significantly different from an identity matrix and whether the correlation coefficients are all zero. The KMO test produces a statistic from 0-1 and indicates the degree that each variable can predict the other. Watkins (2018) suggests setting a value threshold of >0.5 on the KMO test before performing dimensionality reduction.

The PCA has been shown and investigated within the realm of sport, specifically in fatigue monitoring, performance metrics, and talent identification. For example, Aragon-Vargas & Melisa Gross (1997) explored factors that differentiated good jumpers from poor ones, they found that peak and average mechanical power were important predictors in performance. Charoenpanich et al. (2013) looked at the jumping relating to volleyball and found higher erector spinae muscle activation compared to sedentary controls. Floria et al. (2018) explored not the muscles but rather the waveforms of jumps after a 6-week intervention. They found a much deeper and stronger eccentric phase within the intervention groups. Since jumping variations are most commonly performed on force plates, James et al. (2019) reduced the high volume of metrics down to a fewer amount such as Concentric Mean Power/BM, Concentric RPD, Flight Time: Contraction Time, Jump Height, and Peak Power/BM. These findings are similar to Merrigan et al. (2021) who grouped CMJ into the groups of: explosive transfer to concentric power, powerful eccentric loading, CMJ strategy, and Jump Height/Power.

Within the context of workload, Heishman et al. (2020) explored how CMJ variables can be leveraged to monitor fatigue over a pre-season period in basketball. The authors found that although jump height remained the same regardless of workload, RSI-mod decreased as workloads increased. This study demonstrates how fatigue may be monitored indirectly when the correct variables are teased out. RSI-mod has been shown to track how a player jump strategy relates to their explosiveness. Thus, having a lower RSI-mod would indicate lower explosiveness and increased fatigue even though jump height may remain the same. Saucier et al. (2021) used sEMG embedded shorts to come up with 13 metrics of interest and conduct a PCA on the data. They found number of total accelerations, high accelerations, number of total decelerations, and high accelerations accounted for most of the variance within the two principal components in basketball positions. Stone et al. (2022) investigated external workload metrics within the same collegiate basketball population. The authors looked at mechanical workload, jump loads, accelerations, decelerations, average speed, and max speed. They found that total decelerations and max speed were of interest to monitor throughout the season.

One last aspect relates to use of PCA in talent identification since the extrinsic and intrinsic variables related to human performance are vast. McCormack et al. (2021) found that max speed, Peak Force in the IMTP, and body mass were important variables in talent identification. These findings were supplemented by Parmar et al (2018) who approached the question via a tactical standpoint and found amount of possession time and making quick ground which is a tactical metric comprised of workload related to an increased speed of play (maximal velocity and % of max speed) made a team more successful.

Thus, there is information into how the PCA technique can be used for practitioners and clinicians to better understand large sets of data in their decision-making processes. However, there are still gaps in the literature regarding methods, population, and specificity of research. For example, to the authors knowledge research has never been done on the NBA regular season which brings increased game load (82 regular season) with more travel density all over the country rather than in regional conferences as is typically done in college. Furthermore, the financial implications at the professional level (NBA) are higher than at the collegiate level due to higher revenue streams and each player representing a much larger financial investment. Lastly, the restrictions on wearable technology set forth by the CBA make tracking during NBA games more difficult. Currently, Second Spectrum serves as the official optical tracking and game tracking provider for the NBA. As such, Second Spectrum tracks player data and then transports all data into a cloud-based system that is available to each team league wide. As previously discussed, each report has access to 134 metrics for each player in each game. These are more metrics than currently investigated in the collegiate setting since these authors used the travel IMU system rather than the permanent system which is set up around the game arena or practice facility. The travel IMU system can thus only provide a fraction of the full system since it is limited to one central receiver rather than a 360-degree field of play like the permanent system offers. The population of professional basketball players have no specific metrics to look at when it comes to workload and fatigue monitoring. Since workload management has become such a pertinent topic within the NBA, it becomes even more important for teams to be able to discern important metrics from within external workload data. This will allow clinicians and practitioners to make better decisions when it comes to training, recovery, and on-court performance. Thus, applying the PCA to external workload data throughout the full NBA regular season to gain better insights into the most important performance metrics will be the focus of this project.

CHAPTER THREE: METHODS

Overview

Chapter 3 provides the methodology of the research project along with the technical specifications of the PCA technique. The design was identified as a quantitative research project working with a professional basketball sample and applied a dimensionality reduction technique. The primary research question sought to investigate which variables within external workload explained the most variance within the Kinexon LPS. Secondarily, it identified how the PCA technique may be applied specifically regarding principal components and which variables were contained within them. The hypothesis was described with the information from a regular NBA season as fit for dimensionality reduction and how a small number of principal components explained a majority of total variance in the dataset. The participants and settings were described for the project in regard to the regular NBA season along with how the data was collected and analyzed. The technical aspects of the PCA and its application in SPSS were explained in detail. The Chapter concludes with how the data was analyzed and represented within the SPSS package.

Design

This project was a quantitative study that applied a dimensionality reduction technique to retrospective data captured throughout the length of the 2022-2023 NBA regular season. A quantitative approach was chosen to interrogate the multitude of variables output by a typical Kinexon report. All player data that could be identified (names) were deidentified in compliance with league, organizational, and academic research standards to protect subject privacy. A retrospective analysis method was chosen to make data aggregation easier. For example, it was more feasible and practical to wait until the 82-game season had concluded and all data was uploaded into the Kinexon Cloud rather than performing a game-by-game analysis. Also, this approach allowed the investigator to see how many games would be part of the final analysis and decide whether to include any post-season competition (play-in vs play-off games). A descriptive portion was important to highlight the novelty of the population studied, since at the time of this writing workload information had not been investigated in NBA basketball players. The quantitative data (physical performance variables) used in this study were a combination of physical measurement metrics and projections of data. The measurements captured were intrinsically part of the game data since each player performs a numerical amount of work and work events per game. For example, a player may run 1.5 miles and perform 38 sprints throughout a game and is all captured via the Second Spectrum optical tracking system. Thus, no additional time or physical burdens were imposed on players since these data were already being collected via second spectrum. Similarly, using a PCA, the investigator found which variables were most related to physical performance through the creation of principal components. All data were collected during the 2022-2023 NBA regular season by the investigator who is a full-time employee of the organization and has the title of "Sports Scientist/Assistant Strength and Conditioning Coach". The principal investigator has the National Strength and Conditioning Association (NSCA) certified strength and conditioning specialist (CSCS) along with certified performance sport scientist (CPSS) certifications which gives him the necessary training to collect workload data and evaluate it from a data analysis standpoint for performance insights. The PCA as a technique was used to reduce the dimensionality of large data sets for ease of interpretation. This was appropriate because each Kinexon report outputs 134 variables per player, per game. In the average NBA game, 8-10 players step onto the court, are tracked by Second Spectrum and subsequently have a Kinexon report. This can make observations and decision-making overwhelming for practitioners and stakeholders. Although correlation, linear regressions, and factor analysis are options in data analysis, the PCA may be a more appropriate tool to extract the most pertinent variables. Using a correlation allows relationships to be explored between two variables, however a causal relationship cannot be established. A linear regression is used to predict the value of a variable based on the value of another variable, however since this dataset is dealing with numerous variables it makes using this technique slow (Bewick et al., 2003). In the mathematical modeling of a linear regression, the error squares are minimized in the vertical direction, while in the PCA error squares are minimized perpendicular to the line of best fit resulting in an orthogonal regression (Gang Su, 2009). This makes a linear regression inappropriate in finding a line of best fit to the data while the PCA seeks to transform the variables into new linear combinations called principal components. Lastly, a factor analysis is statistically similar to the PCA in that both are rooted in multivariate statistics. However, the goal of a factor analysis is finding the latent variables in a data set. The factor analysis focuses on a conceptual understanding of the data set rather than the best numerical interpretation. The PCA focuses on reducing the number of dimensions in a data set. It achieves this by creating principal components and arranging them in descending order of total variance explained from a numerical model standpoint. Using the PCA as the technique of choice was taken from two related studies by Stone at al. (2022) and Saucier et al. (2021). Stone et al. (2022) looked at external workload data from Division 1 Men's Basketball with the Kinexon Travel Unit. The authors used Bartlett's test of Sphericity and KMO measure of sampling adequacy before implementing the PCA analysis. A one-way analysis of variance (ANOVA) was also used to find variables which significantly increased the odds of accurately assigning an athlete to their position group. Cohen's d was used to calculate effect sizes based on those guidelines set forth by Cohen (Cohen, 1992). Saucier et al. (2021) also collected data on 15 Division 1 Men's Basketball players and recorded data from practices and games using Strive shorts. The authors measured internal and external workload and thus used both the Shapiro-Wilk test to check normality and Levene's Test for equality of variance between game and training sessions. Due to the non-parametric nature of the data, a non-parametric bootstrapping method was chosen to compare game and training sessions. A PCA was then used to create principal components within the variables of external load. Using both of these studies as references, along with thresholds for assumptions testing such as Bartlett's test of Sphericity and the KMO measure, the PCA was chosen as an appropriate technique for this data.

Research Question(s)

RQ1: Which external workload variables have 70% or more of their variance explained by the principal components (communalities extraction value 0.7 of greater)?

RQ2: Will greater than 70% of workload variance be explained by 3 or fewer principal components with Eigenvalues greater than or equal to 1.0?

RQ3: If 3 or fewer components explain 70% or more of workload variance-is there a distinct combination of variables having a component matrix loading score of 0.4 or greater that describe the components?

Hypothesis(es)

The null hypotheses for this study are:

$H_0$1: No external workload variables will have 70% or greater of their variance explained by the principal components (communalities extraction value 0.7 or greater).

$H_0 2$: At least 70% of total variance will not be explained by 3 or fewer principal components with Eigenvalues greater than or equal to 1.0.

$H_0 3$: No combination of workload variables will have a component loading score greater than or equal to 0.4 (rotated component matrix loading value 0.4 or greater).

Participants and Setting

This study was performed on 18 professional basketball players playing in the NBA, the most competitive and highest level of basketball in the world. Each subject was part of the 2022-2023 NBA active roster spanning from October 2022-April 2023. For the 2022-2023 season, the NBA roster was comprised of 15 players and two additional 2-way players. A 2-way player is one who is able to play in both the G-League (Developmental League) and NBA. However, these players are only allowed a maximum of 50 games with the parent organization. When taking into consideration in-season trades and free agent signings, the total number of participants included in the study were 18 professional basketball players (15 roster, 2 two-way players, 1 signed free agent). In an effort to better understand the population being studied, descriptive statistics were also used. These measures included age (yrs.), height (in), weight (kg), and years played professionally. Unlike the college setting where the athletic performance goals are rooted in player development, the purpose of the NBA is primarily to win championships while preserving player health, wellness, and career longevity. To put this into perspective, the NCAA Men's College Season lasts from early November to Early April and consists of 25-35 games per season compared to the NBA season of 82 games lasting from Late October to Mid-April. The primary reason for the range in college season length are amount of pre-season games and post-season advancement in the 'March Madness' tournament. The average college basketball training week consists of individual court sessions 5-6 days per week, weight room 5-6 days per week (30-60 minutes), and team practice 4-5 days per week (2-3 hours) (Edwards et al., 2018). This is contrasted by an average NBA training day which consists of court sessions 2-3x/week, weight room 2-3x/week, and team practice 3-4 days per week (1-2 hours). The decrease in practice and training volume is mainly due to the increased game frequency, game density, and travel. This study only included game data since it serves as the primary tool from a decision-making standpoint in preparation for games. The sampling procedure used was convenience, due to the population being worked with. Convenience sampling is defined as subjects being included in the sample because they are the ‘easiest’ for the researcher to access (Sedgwick, 2013). The principal investigator is a sports scientist/strength and conditioning coach for the organization being studied. While it may have been more significant to sample the entire NBA population, this becomes difficult due to the competition between organizations. Furthermore, obtaining permission from 30 teams plus the NBA league office becomes a logistical challenge. The elite status of NBA athletes and the opportunity to study them present a unique research project, making convenience sampling feasible and appropriate.

SAMPLING

In regard to sampling, guidance was taken from Svilar et al. (2018) who used 13 elite-level (First Spanish Division) athletes and monitored their training sessions only using Catapult Innovations S5 devices (Melbourne, Australia) to gather 300 total observations. In a similar study by Svilar et al. (2018), the authors investigated key metrics for load monitoring using 13 professional basketball players in the first Spanish division (Liga Endesa). This study used data from 18 roster players which was an appropriate sample size as validated by previous research (Stone et al., 2022; Svilar, Castellano, & Jukić, 2018). This project followed the procedures of the aforementioned studies by categorizing the subjects into their respective position groups (Guards n = 7, Forwards n = 7, and Centers n = 4).

Instrumentation

The PCA was the technique of choice used in this study and has been validated as a multi-variate statistical model. Rao (1964) was among the first to explore this technique because it allowed for large data sets to be condensed into principal components while preserving accuracy and data integrity (Rao, 1964.). Since then, the idea of dimensionality reduction has been investigated by various authors (Bro et al., 2014.; Dascalu et al., 2017; Holland, 2019; Jollife & Cadima, 2016; Lever et al., 2017; Richardson, 2009; Jolliffe, 1990) . The PCA has been found to be a valid method to reduce the number of variables in large data sets so long as sphericity and sampling adequacy are assessed. From here, several authors have applied the PCA to the collegiate basketball setting in relation to external workload data (Stone et al., 2022; Saucier et al., 2021) to make contributions into the field of workload monitoring. Svilar et al. (2018, 2018a) was the first to use the PCA to investigate workload data reduction within European professional basketball players (Spanish First Division). Since the PCA has been established as a means of reducing dimensionality within external workload data at the Division 1 and professional level (Europe), this technique was chosen to run the first investigation into professional basketball players in the NBA.

The principal component analysis is a technique rooted in multi-variate statistics and used for dimensionality reduction. Its purpose is to create principal components which explain total data variance in a smaller number of variables than the original data set. This is first achieved through checking sphericity and sampling adequacy. Specifically, Bartlett's Test of Sphericity compares an observed correlation matrix to an identity matrix (Tobias & Carlson, 1969). This is done to check redundancy between variables because if variables aren't related in some way then the data can't be condensed. The significance value set for sphericity should result in less than 0.05 to proceed. The KMO Test for Sampling Adequacy is performed for each variable in the model and total model. The test is a proportion of variance among variables, the higher this value the more suited the data for a PCA (Dziuban & Shirkey, 1974b). This value usually needs to be higher than 0.70 for the PCA to proceed. Once both assumptions have been tested and met, the data set is ready for PCA procedures. This study utilized SPSS for the analysis since it has been validated within the literature and the coursework taken by the investigator at their academic institution used this program (Liu et al., 2003).

Procedures

Since each team/organization owns their own data, IRB approval was obtained from the Chicago Bulls Basketball Organization and from Liberty University before beginning data collection. This study was retrospective in nature and part of the standard NBA game collection so no additional procedures were imposed on players. Due to the retrospective nature of this study, consent forms were not needed from each athlete, however IRB approval was gathered from both the Chicago Bulls and Liberty University. All active roster players were included in this study. The current protocol of this study was to download all game data from the 2022-2023 season from Kinexon into a secure Excel spreadsheet that was only accessible by the principal investigator and committee members.

After each game, the NBA used Second Spectrum to upload data into their cloud storage system upon which Kinexon takes the data and translates it into their proprietary methods to measure player workload via multiple variables. The final Kinexon data was then available to each organization at 5AM EST the following day for review and analysis. This Kinexon data was downloaded as a CSV file into Excel for data aggregation and cleaning. Microsoft Excel (Microsoft Corporation, 2018) was used to aggregate all data from each player through each game of the 2022-2023 season. As guided by Saucier et al. (2021), player information was discarded if they did not accumulate 5 minutes of court time and used both training and game sessions while Stone et al. (2022) did not dismiss any data points and only investigated game data. Although Kinexon data consists of 134 variables, Stone et al. (2022) used 6 variables of interest while Saucier et al. (2021) used 10 external load variables and 6 internal load variables. To further expand on previous research, this study used the full number of available metrics which could be categorized into Basic, External Load, Internal Load, Performance Events, Scouting Events, and Other. However, to reduce data redundancy, 134 variables were interrogated against each other to synthesize the final number (n = 15). The final Excel file was then imported into SPSS (Premium Version 8) for analysis and is described in the following section.

The Kinexon report yields 134 variables per session and many of these variables overlap, making a PCA unfeasible with raw data. With overlapping variables or those derived from each other, the chances for aggregating similar variables incorrectly into their own principal components such as ‘acceleration zone 1’ and ‘acceleration zone 2’ increase. Thus, the author used the following techniques to compile and clean the data before analysis.

The raw data were downloaded from the Kinexon cloud system to Excel for each of the 18 players and each game they participated for a total of 33,320 data points to be investigated. The first step after downloading the data to Excel was to import all variables to SPSS from Excel and ensure their correct format (scale, ordinal, nominal) of which all variables used in the analysis were scale (numeric). Since the permanent Kinexon system and Second Spectrum optimal tracking systems are different (sensor vs optical tracking), descriptive analytics such as age (years), height (in.), weight (kg.), and years as a professional were used to explore the data set for missing values which could hinder the analysis. Implausible values which were either identified as missing or null were then removed from the working data set. All variables were kept track of for documentation purposes from the original amount of variables (134) to final variables (15) and further illustrated in Table 1. Once implausible values were removed, a correlation matrix of all variables was created to assess strength of relationship between each other. A threshold of 0.9 or higher was used to assess a 'high' amount of correlation and further remove variables. For example, 'High Metabolic Power Distance' and 'High Speed and Acceleration Distance' were highly correlated. 'High Metabolic Power Distance' was chosen since it is used more widely in the literature (Stone et al., 2022; Svilar, Castellano, Jukic, et al., 2018). Variables were then removed based on the correlation matrix along with the results of Bartlett's Test of Sphericity and KMO Measure of Sampling Adequacy. Interestingly enough, the two aforementioned measures were missing from iterations of the early analyses due to 'not-positive definite' outputs. The most common cause of this error message within SPSS is either having too few cases or highly correlated variables within the matrix. This process was repeated until all non-redundant variables were identified within each Kinexon performance category (Basic, External Load, Internal Load, Performance, Scouting Events, and Others as possible). In more detail, 134 variables were input into SPSS to reflect a full Kinexon report and reduce subjectivity bias of which variables were the most 'important' without proper statistical analysis. The primary correlation matrix revealed 44 variables had missing values and were excluded from further analysis. These variables included measures of heart rate since although it is possible for

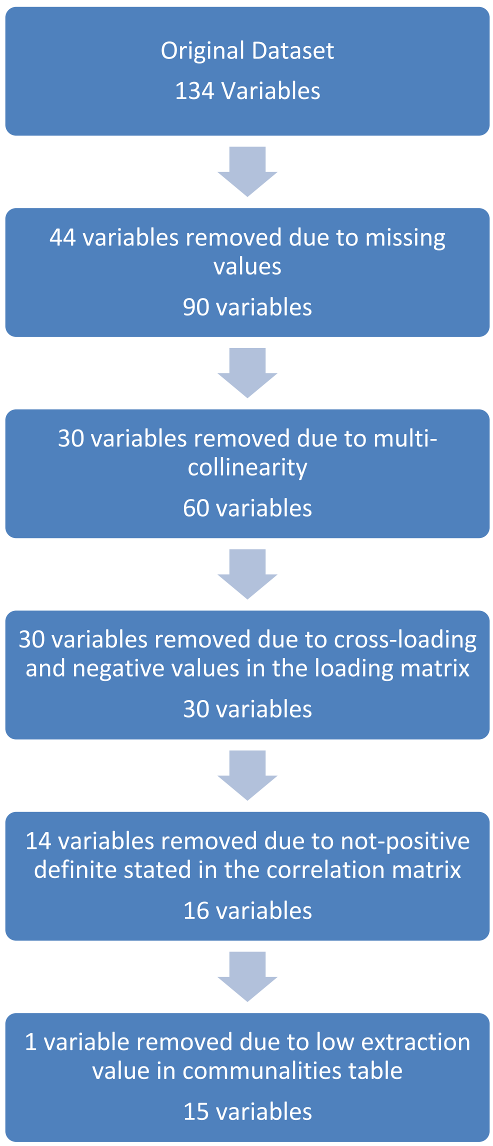
Kinexon to be integrated with a HR monitor, this does not occur in games due to the previously mentioned CBA restrictions on wearable technology. Of the remaining 90 variables, 30 were removed due to a high level of multi-collinearity (threshold set to 0.9). These variables included highly related metrics such as 'Jump Load' vs 'Jump Load per Min' vs 'Jump Load per Mass' which can obscure the final output by putting related variables into a principal component of its own. Of the 60 remaining variables, 30 more variables were removed due to cross-loading and negative values in the loading matrix. Cross-loading indicates correlation on more than one principal component and is indicative of influence by multiple variables. Negative values indicate a negative correlation to the research questions about variables related to external workload and were discarded from further analysis. Of the 30 remaining variables, 14 were removed due to a not-positive definite (NPD) status in the correlation matrix. The NPD error usually stems from either not having enough cases or having variables that are highly related to each other (Wothke, 1993). There were 831 total cases which exceeds the minimum of 150 case set by Shaukat et al. (2016). Thus, more variables with high degrees of correlation were removed from analysis. Sixteen variables remained and a primary PCA was conducted which revealed one variable had a 0.6 extraction value in the communalities table leading to it being removed. Thus, this led to 15 variables being included into the final principal component analysis. This allowed for the output to be created and penultimate data analysis to be further analyzed. The data output another correlation matrix which served as a cross-reference to ensure variables were not highly correlated with each other before final analysis. The communalities table now revealed the extraction factor of all variables crossed the minimum threshold value of 0.7. The total variance explained table was used to extract the initial eigenvalues, percent of variance, and cumulative variance using both an extraction sums of squared loadings and rotation sums of squared loading method. The rotated component matrix via Varimax rotation method with Kaiser Normalization was used to assess for cross-loadings and value of said loading within each principal component. The results of the analysis will be explained in the following chapter.

Data Analysis

The investigator used SPSS (IBM SPSS Statistics for Macintosh, Version 28.0. Armonk, NY: IBM Corp) for the data analysis. The Excel sheet containing all player metrics was uploaded into SPSS for the PCA. In SPSS, the PCA is categorized under 'Dimension Reduction'. As described in Laerd Statistics, the five assumptions made included: the dataset is composed of continuous variables, there is a linear relationship between all variables, there should be sampling adequacy (large enough sample size), the data should be suitable for data reduction, and there should be no outliers (SPSS Statistics Tutorials and Statistical Guides | Laerd Statistics, 2023). The data were analyzed under 'Dimension Reduction' and 'Factor' to begin the analysis. All variables were transferred into the 'variable' column where a correlation matrix, KMO measure of sampling adequacy, and Bartlett's test of sphericity were selected. Under 'Extraction' the method for analysis was 'Principal Components' with correlation matrix, the display was 'Scree plot' and the Extract was based on Eigenvalues great than '1.0'. The rotation of the model used was 'Varimax', and the data were displayed in a 'Rotated Solution' and 'Loading Plot' with '25' selected as the maximum iterations for convergence. The model used a 'Regression' method to save as variables. Lastly, missing values were handled by excluding cases using a listwise manner. The coefficient display format was formatted by size and suppress coefficients smaller than 0.35 for ease of interpretation(Bryant & Yarnold, 1995). The output showed a correlation matrix, KMO and Bartlett's Test, total variance explained, scree plot, rotated component matrix, and component plot in rotated space which were expanded upon in Chapter 4.

Figure 1

Flowchart describing the process of variable selection.



CHAPTER FOUR: FINDINGS

Overview

This chapter provides a description of the sample studied in the project which consisted of 18 NBA basketball players. Descriptive statistics such as age and years playing professionally were used to orient the reader into their athletic profiles since research on professional basketball players is scarce. The findings of the principal component analysis (PCA) were then presented in a detailed manner where readers may understand and replicate its parameters. Whereas most research into dimensionality reduction does not describe how the variables were obtained, this chapter will explain each step of the PCA the requirements met to proceed with every step of the final analysis.

Research Question(s)

RQ1: Which external workload variables have 70% or more of their variance explained by the principal components (communalities extraction value 0.7 of greater)?

RQ2: Will greater than 70% of workload variance be explained by 3 or fewer principal components with Eigenvalues greater than or equal to 1.0?

RQ3: If 3 or fewer components explain 70% or more of workload variance-is there a distinct combination of variables having a component matrix loading score of 0.4 or greater that describe the components?

Null Hypothesis(es)

$H_0 1$: No external workload variables will have 70% or greater of their variance explained by the principal components (communalities extraction value 0.7 or greater).

$H_0 2$: At least 70% of total variance will not be explained by 3 or fewer principal components with Eigenvalues greater than or equal to 1.0.

$H_0 3$: No combination of workload variables will have a component loading score greater than or equal to 0.4 (rotated component matrix loading value 0.4 or greater).

Descriptive Statistics

Since the NBA is the highest level of professional basketball in the world, achieving this level of play is extremely rare not only from an athletic standpoint, but also from a research standpoint. Thus, this project provided a unique opportunity to further understand basketball sports performance at the highest level. In the US alone, 27.14 million people played basketball recreationally in 2021 (U.S. Americans Who Played Basketball 2021 | Statista, 2023). Within more competitive basketball, only 2.9% of boy's high school basketball players will get the opportunity to play collegiate basketball at any level. Only 1.3% of those collegiate athletes (primarily Division 1) will get the opportunity to play professional basketball. Thus, being able to study 18 of the 450 current NBA players presented a unique opportunity.

The subjects from this study all played for the same NBA organization in the Central division of the Eastern conference. The NBA consists of 30 teams, split into 15 teams competing in two conferences (Eastern, Western), and further divided into 6 divisions (3 in the East, 3 in the West). A total of 18 players (n=18) were included in this study due to trades and free agency since only 15 active players may be on the roster at any point. The average age of the subjects was 27.2 years (± 4.77). The average height of subjects was 78.2 inches (± 3.20) and average weight 97.1 kg (± 12.84). The average length of professional career was 7.2 years (± 4.51) (Table 1). Since there are 82 games in one NBA season, this study began with 8938 data points per player and 160884 total data points for all subjects (n=18) which was reduced to 831 data points in the final analysis (Principal Component Analysis) (Table 1).

Table 1

Descriptive Statistics of the NBA athletes studied.

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Years Professional	7.20	4.51	831
Weight (kg.)	97.10	12.84	
Height (in.)	78.20	3.20	
Age (years)	27.20	4.77	
Distance (mi)	1.61	.74	
Speed (% of max.)	77.40	8.37	
Acceleration (max.) (ft/s²)	11.23	1.41	
Accumulated Acceleration Load	313.92	142.08	
Decels Load	619.07	283.88	
High Metabolic Power Distance (mi)	.27	.12	
Mechanical Load	826.54	378.51	
Metabolic Power (max.) (W)	3695.82	2920.82	
Metabolic Work (kcal)	59.84	103.88	
Accelerations	81.65	37.05	
Decelerations	73.45	33.31	
Sprints	36.30	16.60	
Time (Sec)	1418.97	672.37	
Met Power Time High (Sec)	141.64	65.19	
Speed (max.) (mph)	14.72	1.59	

Table 2

KMO and Bartlett's Test

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.877
Bartlett's Test of Sphericity	Approx. Chi-Square	42522.95
	df	105
	Sig.	<.001

Each Kinexon report began with 134 variables and were reduced into 15 final variables. Bartlett's test assesses whether a correlation matrix is significantly different from an identity matrix. This test checks for correlation between at least some of the variables in the data. The KMO statistic is a value between 0-1 and indicates whether the sum of partial correlations is large relative to sum of correlation. A value of 0.8 to 1.0 indicate the sampling is adequate, 0.7 to 0.79 are middling, and 0.6 to 0.69 are mediocre. Values less than 0.6 are usually not suitable for dimensionality reduction unless remedial steps are taken (Shrestha, 2021a).

Results

Hypothesis(es)

Null Hypothesis 1: $H_0$1: No external workload variables will have 70% or greater of their variance explained by the principal components (communalities extraction value 0.7 or greater).

For this null hypothesis, the principal component was performed only when the two assumptions of KMO sampling adequacy and Bartlett's test of sphericity were met. The KMO sampling adequacy was .877 which is adequate (refer to Table 2) (Shrestha, 2021b). Bartlett's

Test of Sphericity yielded a significance value of <0.001 (refer to Table 2) which meant the null hypothesis was rejected. In this application, this signified the data was not an identity matrix and there was sufficient correlation between all variables (Shrestha, 2021b). Since these assumptions were met, the PCA was continued and performed successfully. The final correlation matrix was positive definite (there were no negative values). Highly correlated variables (>0.9) were excluded to reduce multi-collinearity in the final model according to the paper by Shrestha (2020). To explore variables of importance from the final model, the communalities table was investigated for each variable and its extraction value. The final 15 variables are listed in Table 3 of which the variables had an extraction factor higher than 0.7 and are listed in descending order. Accumulated Acceleration Load (.989), Mechanical Load (.985), Distance (mi) (.985), Decels Load (.984), Accelerations (.975), Met Power Time High (.969), Time on court (.959), Decelerations (.955), Max Speed (.951), Speed (% of max) (.951), and High Metabolic Power Distance (.921), Sprints (.870), Metabolic Work kcal (.783), Acceleration Max ft/s^2 (.756), Metabolic Power Max (.726). The findings from the communalities table indicate there are identified variables in the final model as described by their extraction values greater than 0.7 which caused the author to reject this null hypothesis.

Table 3

Communalities Tables

Communalities

	Initial	Extraction
Accumulated Acceleration Load	1.000	.989
Mechanical Load	1.000	.985
Distance (mi)	1.000	.985
Decels Load	1.000	.984
Accelerations	1.000	.975
Met Power Time High_(Sec)	1.000	.969
Time (Sec)	1.000	.959
Decelerations	1.000	.955
Speed (max.) (mph)	1.000	.951
Speed (% of max.)	1.000	.951
High Metabolic Power Distance (mi)	1.000	.921
Sprints	1.000	.870
Metabolic Work (kcal)	1.000	.783
Acceleration (max.) (ft/s²)	1.000	.756
Metabolic Power (max.) (W)	1.000	.726

Extraction Method: Principal Component Analysis.

$H_0$2: At least 70% of total variance will not be explained by 3 or fewer principal components with Eigenvalues greater than or equal to 1.0.

From the 134 Kinexon variables that the original dataset contained, 16 variables were included in the analysis and then further reduced to 15 variables (Figure 1). These variables were synthesized using correlation matrices, checking for multi-collinearity, and cross-loading. Since 15 variables were used in the final analysis, there were 15 principal components created. Using the Total Variances Explained (Table 4), 3 principal components stood out and explained 91.7% of the total variance. The 3 principal components explain the following amount of variance: Principal component 1 (individual 68.9%, cumulative 68.9%); principal component 2 (individual 14.1%, cumulative 83.1%); principal component 3 (individual 8.6%, cumulative 91.7%). Previous research has suggested total variance in a model should cross the threshold of 70% which this model indeed crossed (Cangelosi & Goriely, 2007). The Eigenvalues for each principal component were as follows: Principal component 1 (10.34), principal component 2 (2.12), principal component 3 (1.28). The scree plot (Figure 2) also supported this finding as the first three principal components are above the inflection point of the curve and have eigenvalues greater than 1.0 Since a majority of the variance (91%) was explained by 3 principal component and have Eigenvalues greater than 1.0, the second null hypothesis was rejected.

Table 4

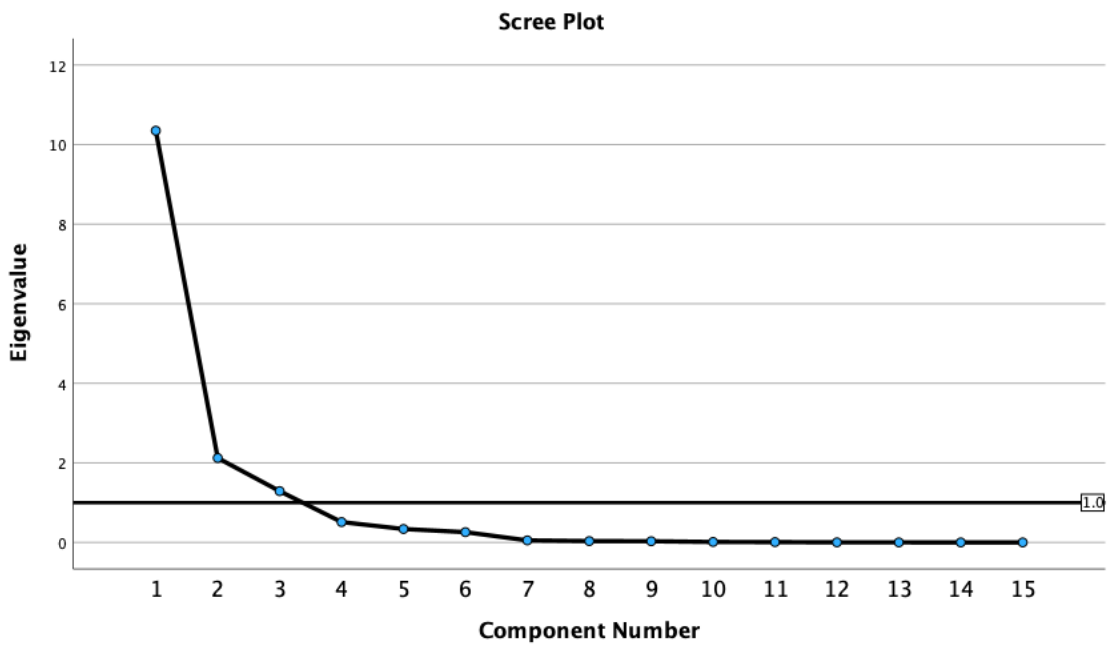
Total Variance Explained

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.34	68.99	68.99	10.34	68.99	68.99
2	2.12	14.14	83.13	2.12	14.14	83.13
3	1.28	8.59	91.72	1.28	8.59	91.72
4	.51	3.40	95.13			
5	.33	2.24	97.38			
6	.25	1.72	99.10			
7	.05	.33	99.44			
8	.03	.22	99.66			
9	.02	.19	99.85			
10	.01	.074	99.93			
11	.008	.052	99.98			
12	.002	.011	99.99			
13	.001	.004	99.99			
14	.000	.001	100.00			
15	6.407E-5	.000	100.00			

Figure 2

Scree Plot



$H_0$3: No combination of workload variables will have a component loading score greater than or equal to 0.4 (rotated component matrix loading value 0.4 or greater).

Within the principal component analysis, two tables can be looked at for principal components and their respective loading scores. The component matrix (Table 5) displays component loadings for each item prior to rotation (Rotation = 'None'). The rotated component matrix (Table 6) was the key output because it contains the component loadings after rotation (Rotation = 'Varimax' with Kaiser normalization). An orthogonal rotation is necessary to obtain the 'optimal' angle on lines that will run through the most amount of data and subsequent variance. Although there are different types of rotational options when performing a PCA, the varimax option maximizes the sum of variance of the squared loadings. This rotation allows for a smaller subset of more defined variables which seems to be the point of dimensionality reduction (Abdi, n.d.). The rotated component matrix was compared to the original component matrix to compare for extraction values and ensure cross-loaded variables were not present. Three principle components were present in the rotated component matrix, were sorted by size, and values less than 0.35 were suppressed for increased clarity of the final model. Principal component 1 contained: Mechanical Load (.976), Decelerations Load (.976), Distance (.971), Time (.970), Accumulated Acceleration Load (.968), Met Power Time High (.968), Decelerations (.949), Accelerations (.945), High Metabolic Power Distance (.897), Sprints (.880). Principal component 2 contained: Speed % of Max (.944), Max Speed (.943), Max Acceleration (ft/s^2) (.831). Principal Component 3 contained Metabolic Work (.873), Metabolic Power (.812). Of all variables, all loading values were above 0.8 which was higher than the set threshold of 0.4. From a visual standpoint, the component plot in rotated space allows a clearer presentation between each variable and their respective groupings into principal components. It is possible to visualize all 15 variables and 3 principal components into one plot of 3-dimensional space (X-Y-Z axes), but this makes interpretation difficult since information is too cluttered. Since 3 principal components were produced from the principal component analysis (Eigenvalue > 1.0) 3 component plots were created. Each component plot contained two principal components and their respective variables (Figure 3, 4, 5). Figure 3 contained principal components 1 and 2, Figure 4 contained principal components 1 and 3, Figure 5 contained principal components 2 and 3. Each respective figure confirms how component 3 shows a grouping between Metabolic Work (.871) and Metabolic Power (.812). Component 2 shows a grouping between Max Speed (.943), Speed % of Max (.944), and Acceleration Max (.831).

Component 1 shows groupings of variables such as Mechanical Load (.976), Accumulated Accelerated Load (.968), and Accelerations (.945). Within each component plot, this can be confirmed visually through the groupings of variables pertaining to each principal component. Since each of the 15 variables contained in each of the three principal components had extraction values higher than 0.4 and no cross-loading between variables, this null hypothesis was rejected.

Table 5

Component Matrix

Component Matrix[a]

	Component		
	1	2	3
Accumulated Acceleration Load	.984		
Accelerations	.983		
Distance (mi)	.978		
Mechanical Load	.974		
Decels Load	.973		
Decelerations	.968		
Met Power Time High (Sec)	.964		
High Metabolic Power Distance (mi)	.959		
Time (Sec)	.949		
Sprints	.932		
Speed (% of max.)	.541	.773	
Speed (max.) (mph)	.542	.772	
Acceleration (max.) (ft/s²)	.503	.689	
Metabolic Work (kcal)		.411	.767
Metabolic Power (max.) (W)	.377		.733

Extraction Method: Principal Component Analysis.
a. 3 components extracted.

Table 6

Rotated Component Matrix

Rotated Component Matrix[a]

	Component		
	1	2	3
Decels Load	.976		
Mechanical Load	.976		
Distance (mi)	.971		
Time (Sec)	.970		
Accumulated Acceleration Load	.968		
Met Power Time High (Sec)	.968		
Decelerations	.949		
Accelerations	.945		
High Metabolic Power Distance (mi)	.897		
Sprints	.880		
Speed (% of max.)		.944	
Speed (max.) (mph)		.943	
Acceleration (max.) (ft/s²)		.831	
Metabolic Work (kcal)			.871
Metabolic Power (max.) (W)			.812

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Figure 3

Component Plot in Rotated Space (1,2)

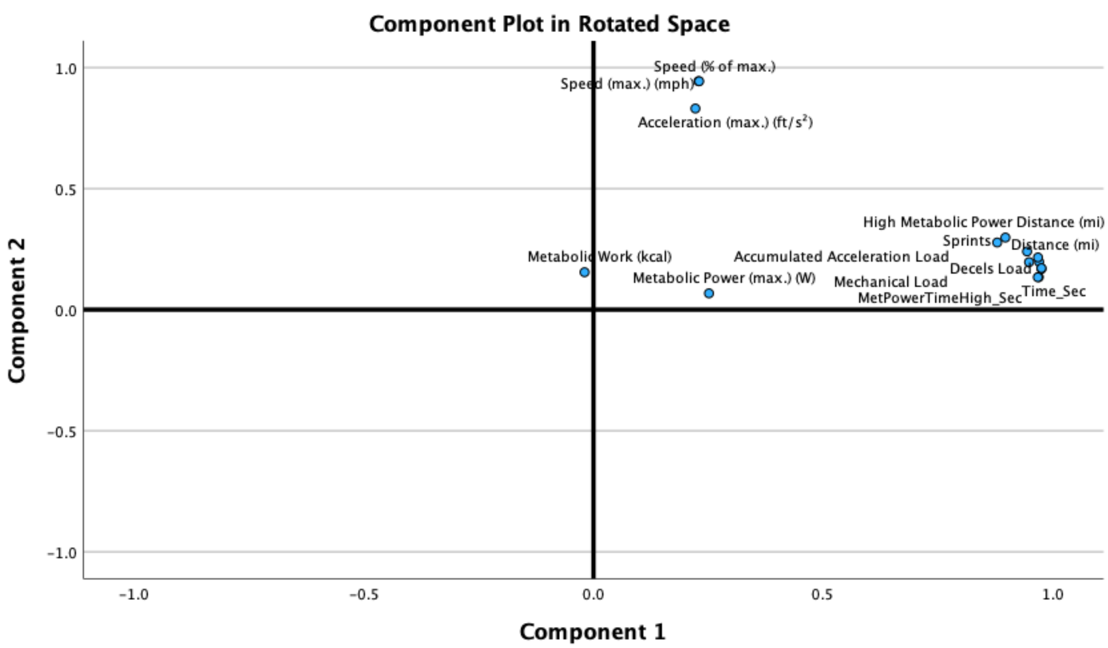


Figure 4

Component Plot in Rotated Space (1,3)

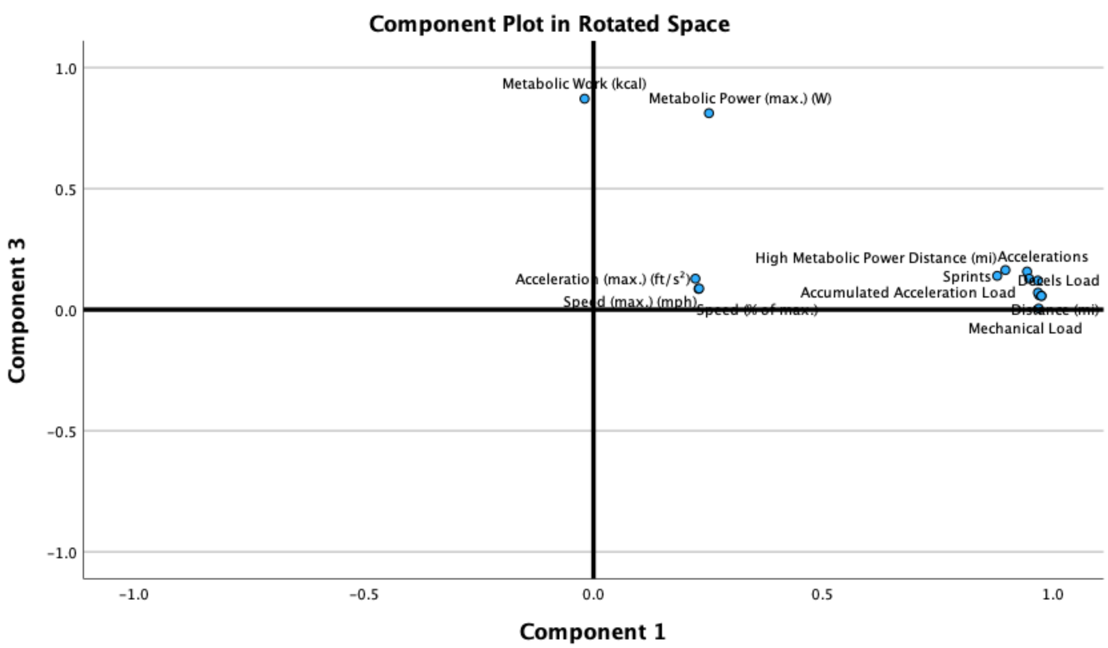
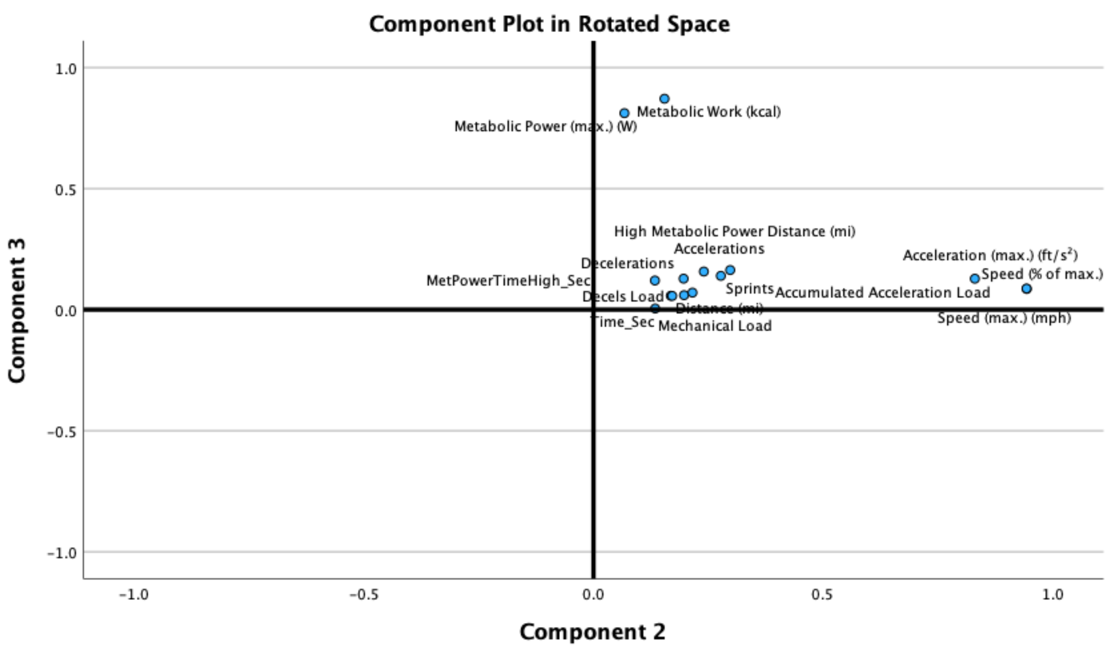


Figure 5

Component Plot in Rotated Space (2,3)



CHAPTER FIVE: CONCLUSIONS

Overview

This chapter will begin with a discussion of how the principal component analysis was conducted and applied to the sample of NBA basketball players. This study sought to answer three primary research questions regarding variables of importance within the Kinexon LPS system of which each one will be described in detail. The discussion portion will focus on comparing the findings from the present study to previous literature, specifically why discrepancies may have been found and how methods differed. Implications of this study will be discussed within the context of NBA players and this unique population. Limitations will be covered, for example how data was only collected over the course of one competitive season. All of these factors will then lay the foundation for creating the section to recommend future research to keep the field of sports science progressing.

Discussion

The purpose of this study was to explore the most relevant variables within the Kinexon LPS using a multi-variate statistical (principal component analysis) approach. Since the principal component analysis has multiple steps, the primary research questions revolved around finding the most relevant variables from the communalities table, have 3 or fewer principal components, and relevant variables from the rotated component matrix. Each research question was answered and will be discussed in further detail within the following section.

Historically, the process of data collection, analysis, and dissemination has been placed on the strength and conditioning coach. This is complicated because the field of data analytics within sport has turned into sports science which is still poorly defined. This gap in practice makes providing the highest level of care difficult because the strength and conditioning coach or sports medicine clinician has been pushed to learn an entire new field in addition to their daily duties. Recently, Gleason et al. (2023, pg. 2) defined sports science as “the field in which multiple scientific disciplines are studied and applied concurrently to understand and improve sport performance in an interdisciplinary or transdisciplinary manner through observation, management, and modification of athlete selection, development, training, and recovery processes, optimization of equipment, and detailed analysis of technique and tactics; this is optimally accomplished through a focus on the athlete, careful coordination with sport and strength and conditioning coaches, and the overt support of coaches and administrators”. Within the context of this study, the primary investigator serves as a sports scientist/ strength and conditioning coach for an NBA organization. The definition from Gleason et al. (2023) would apply in the management and modification of training to optimize on-court performance. Since the NBA and its technology have evolved so much within the last 20 years, LPS systems such as Kinexon have come to market as ways to quantify external workload. By being able to quantify external workload during games and training sessions, performance health staffs are better able to manage athletes and their performance throughout the 82-game season. According to Gleason et al. (2023, pg. 3), the sports scientist “provides expert advice and support to athletes and coaches to help them understand and enhance sport performance; adopting evidence-based, quality-assured practices to evaluate and develop effective strategies or interventions in training and competition”. By being able to quantify external workload, the sport scientist can help guide training and recovery interventions. A primary challenge however, is how to sort through the plethora of variable output by a typical Kinexon report which contains 134 variables. Beyond this, the micro- (athlete), meso- (coaching staff), and macro-levels (front office/organization) of stakeholders makes communicating complex information to different levels of knowledge important to success. Using the common specialties and subspecialities proposed by Gleason et al. (2023), the principal investigator leaned into 'sport biomechanics' for this project since Kinexon is a LPS system that utilizes tri-axial accelerometry and optical tracking through second spectrum to obtain external workload. Over the 2022-2023 season, all 82 games were captured for 18 players and then put through a principal component analysis to find the most pertinent variables of external workload to optimize on-court performance as it relates to professional basketball. The primary research questions were as follows:

RQ1: Which external workload variables have 70% or more of their variance explained by the principal components (communalities extraction value 0.7 or greater)?

Saucier et al. (2021) investigated external load and muscle activation in 15 Division 1 players. The authors used a total of 16 external load and muscle load variables in their analysis and justified their selection through previous literature and interest of the coaching staff. From 74 original variables, they narrowed variables down to 15. However, although non-parametric bootstrapping (two sample bootstrap) and PCA were both used to extract principal components and variance, their original variables were not complexly justified. The following articles are used as justification (Russel et al. 2020; Petway et al., 2020; Scanlan et al., 2015; Sampaio et al., 2015; Puente et al., 2017; Svilar et al., 2018). Russell et al. (2020) performed a systematic review on the methods and technologies used to monitor the physical demands in basketball. This article included 122 studies in the final review of which 41 included measures of both internal and external training loads, 29 measured external training load only, and 52 measured internal training load only. The authors found common methods of measuring workload included inertial devices, positional systems, heart-rate derived load, and subjective measures such as session rating of perceived exertion (sRPE). In the discussion, the authors stated how the vast disparities in basketball training load monitoring methodologies lead to inconsistencies in reporting physical demands. Since the principal component analysis used in this study yielded 15 pertinent variables for monitoring external workload in professional basketball, more evidence has been made available to support which variables performance coaches should focus on. Furthermore, these variables were not based on previous literature where the analysis process is not well described or mainly relies on subjective stakeholder input. However, this study was based on a rigorous process through the principal component analysis.

Petway et al. (2020) described basketball as a sport where acceleration, deceleration, change of direction, jumping, and shuffling are important from a tactical and technical standpoint. Thus, the results of the communalities table (Table 3) show max acceleration (.756), accumulated acceleration load (.989), and decelerations (.955) are in line with the thoughts by Petway et al. (2020) with the difference of jumping. Although the connection between basketball and jumping may seem intuitive, this could not be measured since the optical tracking system only tracks linear and lateral movements. Petway et al. (2020) also explained the relevance of total high-intensity accelerations, total distance traveled, and top speed which was also found by the communalities table (Table 3) (Distance [.985], Speed % of max [.951], High Metabolic Power Distance [.921]). A systematic review was conducted on 3,282 total articles and the final inclusion analysis was narrowed down to 35 studies (28 elite competition demands and training load) and found the most important outcomes to be total distance, acceleration, deceleration, average speed, and top speed to be the outcomes extracted. Interestingly, although training load was not measured in this research study, Petway et al. (2020) found accelerations, decelerations, and change of direction to be variables extracted from the systematic review which also aligns with game variables from both studies. An interesting point is that although both studies line up in terms of variables, other than assessing each study using the PEDro scale for scientific quality there were no statistical means of obtaining the variables concluded unlike in the present study. Scanlan et al. (2015) explored the fluctuations in activity demands across game quarters in professional and semi-professional basketball. The authors used a video-based time-motion analysis (TMA) made to capture player movement (n=10) during games-which is similar to the second spectrum optical tracking system used in the NBA, then converted to Kinexon metrics. The metrics measured in the study by Scanlan et al. (2020) were more generalized into: Low-intensity activity, High-intensity activity, shuffles, dribbles, jumps, upper body, and total. The authors found the largest differences were between the first and third quarters in regard to dribbling, total velocities, and high-intensity activity. Although a direct comparison cannot be made between TMA variables and motion-based ones (Second Spectrum), it does show the potential importance of total velocity (max speed) and high-intensity activity (high-metabolic power since) in basketball since there is a significant drop from the first to the third quarter. Although this point may seem obvious in that fatigue causes a drop in performance metrics over the course of a game, knowing which variables decrease and at what rate make this finding valuable. For example, since Scanlan et al. (2020) used 7 different variables it would make sense that fatigue would accumulate from the first to third quarter at an equal rate between all variables. This is not the case however since high intensity activity and total velocity showed the largest decrements. Thus, the study by Scanlan et al. (2020) supported the findings from the present research study in variables even though the authors looked at only two games in the Australian NBL vs 82 games in the NBA.

Sampaio et al. (2015) took an interesting view of player tracking within the NBA by looking at all-star vs non all-star players. The authors describe how sports performance data such as distance, velocity, and acceleration activity workload can be combined with technical data to create a more holistic view of on-court performance. A primary objective of this study was to analyze different performance profiles based on athletic success (all-star vs non all-star). Although the variables of focus were primarily technical (pull-up shots, catch and shoot, close shots); speed and distance were also examined. Of all variables, total distance and average speeds were not discriminant variables while speed on defense was a discriminant variable. The authors concluded that in general all-star players cover less ground and shorter distances at lower average speeds. This may allow them to make more calculated decisions in playing and be more successful with their intentions on the court. This partially supports the findings of the current paper since speed was a pertinent variable; the discrepancy may have come from the comparison between two groups of already extremely gifted professional basketball players being judged on technical performance rather than physical workload.

Puente at al. (2017) also looked at physical and physiological demands of players during basketball games. The authors describe sprinting and jumping as key factors to scoring and being successful in the sport. Puente et al. (2017) used 25 male basketball players in the national basketball league over the course of 4 seasons (2009-2014). The SPI Pro X was used in a harness paired with a Polar heart rate monitor to obtain GPS/accelerometry/ heart rate information. A one-way ANOVA was used to measure differences between variables and each playing position (center, guard, forward) with additional Cohen's effect size to assess the magnitude of differences after normality was assessed via Kolmogorov-Smirnov test. The GPS system used was able to assess five different zones: Zone 1 (standing and walking; <6 km/h), Zone 2 (jogging or low-speed running; from 6.1 to 12 km/h), Zone 3 (running or moderate-speed running; from 12.1 to 18 km/h), Zone 4 (high-speed running; from 18.1 to 24 km/h), and Zone 5 (maximal speed running; > 24 km/h). The authors found that players covered a mean distance of 82.6 m/min, although this tournament used two 10-minute halves rather than NBA length games (four, 12-minute quarters). Players spent 81% of their time at low-intensity running and 3% at a high-intensity speed. Even though only a small amount of the game, the authors explain how repeated sprint ability is a key factor to team performance. Thus, increasing running pace at higher intensities may represent a meaningful advantage for success in basketball. This seems to agree with the importance of maximal speed, percentage of max speed, and high metabolic distance.

RQ2: Will a majority of the variance within external workload be explained by 3 or fewer principal components with Eigenvalues greater than or equal to 1.0?

Charoenpanich et al. (2013) used a PCA to identify the major muscles recruited during a vertical jump in volleyball players and sedentary controls. Although the way in which principal components were arrived at was not mentioned, the first principal component accounted for 55% of total variance. In volleyball players, the vastus medialis, tibialis anterior, and erector spinae were the most used while the medial gastrocnemius played a larger role in the sedentary control group. Although the PCA method was specified in technical detail, the reasoning for only using one principal component was not described. However, using one principal component aligns with this current study in that 89% of all variance was described with 3 principal components.

Specifically, the findings from Charoenpanich et al. (2013) support this paper because a fewer amount of principal components (1) explain a majority of the variance (55%) from the original variables (muscle groups).

Floria et al. (2019) also used a small amount of principal components when looking at vertical jump performance. The authors identify a gap in the literature that traditional data reduction methods convert continuous measures to discrete measures but in the process, they potentially discard important information. This is similar to the literature in current workload management, which is based on variables perceived by coaches to be important and previous literature where specific processes are not detailed. While coaching expertise is an important factor in identifying relevant variables, this study sought to add technical specifications in performing a PCA on external workload information such as the study cited in vertical jump performance. Floria et al. (2019) analyzed unique data in using a PCA to assess potential changes in force, velocity, and displacement-time curves on 34 female basketball players. Floria et al. (2014) used the PCA approach described by Deluzio et al. (2007). The results showed 6 principal components in force, 4 principal components in velocity, and 2 principal components in displacement. Although the amount of principal components ranged from 2-6 in the study, this was due to different types of curves and waveforms being investigated of which the 3 principal components found in the current study are still appropriate in reducing from the original 15 principal components due to 15 variables used in the analysis.

Deluzio et al. (2007) used a principal component analysis to reduce the dimensionality in gait analysis due to its complexity as visualized by the different waveforms (joint angle, joint angle moments, etc.). The PCA was used within gait waveform analysis between two groups: subjects with end-stage knee osteoarthritis and a control group. The authors chose to analyze knee flexion angle, adduction moment, and flexion moment waveforms. Each gait measure waveform used a maximum of 3 principal component while adding a t-test to compare between groups (OA vs control) and stepwise discrimination to separate between both groups. The stepwise discrimination and t-test were added to find difference in gait mechanics between two groups while the current study in basketball only had one total group of players which explains only using the PCA. However, Deluzio et al. (2007) used a limit of 3 principal components just like this study did.

James et al. (2021) also explored dimensionality reduction for countermovement jumps metrics from 3 different cohorts (recreationally trained males, competitive MMA athletes, and junior Australian football athletes). Similar to how Kinexon is usually taken at face value for variables like ‘Mechanical Load’, countermovement jumps also experience this fallacy from variables like ‘Jump Height’. However, being able to understand where these numbers and measurements come from forms a more thorough understanding into athletic performance and readiness status. For example, Morin et al. (2018) looked at measuring lower limb maximal power through the squat jump and countermovement jump. The authors explore that due to push-off distance and its derivatives (limb length and countermovement depth), this measure alone may skew any relationship between maximal power and jump height. Also, the athlete may need to have an external load in order to achieve their maximal power output due to their force-velocity profile. Due to these factors, the authors conclude lower limb power may be better assessed through other methods. Similarly, Heishman et al. (2019) examined RSI-mod through the flight-time or impulse-momentum methods in the CMJ with and without arm swings. The CMJ variables were greater when performed with arm swing overall, and each method was correlated to each other. The authors conclude the RSI-mod calculation gives insight into changes in performance due to its ability to assess an athletes jump whole jump strategy. However, jump height was not mentioned due the variance in athletic jump strategy which may mask true changes in performance. Gamble (2021) examined the validity and reliability and Kinexon in ice hockey players against the 1080 sprint device which measures speed and acceleration. Of all variables measured, variables like peak speed and peak acceleration had low standard error of measurements with high intra-class correlation coefficients. Mechanical Workload did not seem to be included which may be explained due its inherent calculation of multiple metrics into one (Acceleration zones 1-4 + Deceleration zones 1-4) causing potential noise in its output. An interesting aspect of this study is the addition of two different force platforms (Bertec, ForceDecks) and 1 linear position transducer (LTP) (MuscleLab). Although the authors used the PCA on the final analysis, it is unclear how they chose the mentioned variables such as concentric impulse, concentric RFD, peak power/BM, etc. The PCA explained about 90% of total variance in 3-4 principal components. This is backed by the scree plot which shows a point of inflection where the eigenvalues drop below 1.0 after a maximum of 4 principal components. However, it is also unclear if the authors make their final recommendations in terms of important variables from the total variance explained rather than the communalities table. If that is the case, this is not supported in the mathematical literature.

Kollias et al. (2013) used the PCA because it allows for a more specific approach compared to the multiple regression linear work previously done by (Aragon-Vargas & Gross, 1997; Dowling & Vamos, 1993; Hay, Dapena, Wilson, Andrews & Woodward, 1978; Podoslsky, Kaufman, Cahalan, Aleshinsky & Chao, 1990) since whole body models accounted for about 88% of total variance and segmental models account for 60% of total variance. Once again, variables chosen for the analysis were selected based on previous literature (Aragon-Vargas & Gross, 1997;

Dowling & Vamos, 1993) rather than statistical methods as is being done in this study. Since 6 variables were examines, 6 principal components were created of which 2 principal components accounted for 74% of total variance in the squat jump.

Laffaye & Bardy (2007) explored the vertical jump from a unilateral standpoint using the PCA within different sports such as volleyball, basketball, and high jumpers. This work built off Kolias et al. (2013) who found differences in squat jump performances could be explained by two principal components and 73% total variance. During the running one-leg vertical jump, kinetic data was obtained with a force plate (AMTI, Massachusetts, USA) and kinematic data with six motion-capture cameras (VICON, Colorado, USA) and 32 passive reflective markers. The authors found two principal components explained 78.3% of total variance and were kept as final results. When conducting a principal component analysis, the purpose is to condense the amount of original variables into the ones which explain the most variance. In alignment with the work by other authors in the field, 3 principal components were kept in the final analysis since they explained 90% of total variance.

RQ3: If 3 or fewer components explain 70% or more of workload variance-is there a distinct combination of variables having a component matrix loading score of 0.4 or greater that describe the components?

McCormack et al. (2021) used a principal component analysis to identify important physical qualities in youth rugby for future development. Tests included a battery of anthropometrics, muscular power, and strength to a total of 654 subjects. The principal components showed running speed, absolute strength, and body mass were all important qualities for future success in the sport.

Merrigan et al. (2021) used the PCA approach to identify the most important variables among Division 1 football players. Like in Kinexon, many variables of the CMJ overlap making a traditional statistical analysis subject to a high degree of multi-collinearity such as peak velocity and peak power. Similar to the present study, Merrigan et al. (2021) used the correlation matrix rather than a covariance method since variables were of different scales and may cause an increase in variance representation due to size. Variables such as explosive transfer to concentric power, eccentric loading, countermovement strategy, and jump height and power were found to be most important.

Panoutsakopoulos et al. (2014) also used the squat jump as a method to assess differences between sports in augmentation to Kollias et al. (2001) who found athletes use a varying amount of force and temporal parameters to achieve outcomes in a countermovement jump. For example, track and field athletes rely on force-dominant patterns while basketball players use a more temporal based approach. The PCA found 6 variables to be the most important within the squat jump which aligns with the previous work by Kollias et al. (2001).

Parmar et al. (2018) used a total of 10 principal components which explained 81.8% of variance to develop performance indicators in professional rugby. While this number is much more than the 3 principal components used in the present study, the discrepancy may lie in that Parmar had 45 variables while the current author used 15 variables. The current study does align with Robbins et al. (2021) who performed 8 PCA's within the same study to analyze joint angles of the ankle, knee, and hip in each plane of movement (sagittal, frontal, and transverse) to identify waveform characteristics. The authors stated they used the first three principal component since "they often account for the majority of variability" and are supported by the current study (Robbins et al., 2021, pg. 135) . In looking at initial torque and maximal torque between track and field and volleyball athletes, Rousanoglou et al. (2006) were able to reduce 9 variables and principal components down to two while explaining 90% and 85.1% of total variance. Lastly, Welch et al. (2019) conducted a PCA on ground reaction forces (GRF) and cutting ability in performance. The authors also wanted to understand GRF and its relationship to max strength, explosive strength, and cutting performance. Twelve principal components were found during the 45-degree cut and 13 principal components were found within the 110-degree cut. Combined, the principal components accounted for 99% of total variance. It was concluded that during the 45-degree cut, a higher concentric and eccentric horizontal to vertical impulse ratio led to better cutting performance outcome. In the 110-degree cut, the previously mentioned outcome of 250 ms. with 500 ms. vertical RFD led to better cutting performance. The literature seems to support that variables related to muscular strength, muscular power, and speed are all related to being important predictors of performance. This agrees with the current study which found variables within Kinexon dealing with total distance, high metabolic power distance, and max speed are also important variables in external workload management.

Implications

This study was unique because it involved a distinctive population using a statistical approach which has not been done before. Basketball as a sport has been studied from a tactical, technical, and performance standpoint. Within the realm of performance: strength and conditioning, programming, and performance outcomes have also been studied. More specifically, since this paper focused on external workload and the sport, tools like isometric strength testing, force production, and training via LPS can and have been studied as well. Isometric strength tools can evaluate isolated strength like the Norbord (VALD, Queensland, Australia) for hamstring strength, Force Frame (VALD, Queensland, Australia) for hip abduction and hip adduction among other muscle groups, and isometric mid-thigh pull for a more global measure of strength and rate of force development. Force production can be tested many ways but force plates (Foredecks, VALD, Queensland, Australia) seem to be popular in sport due to their portability, affordability, and ability to capture a multitude of variables in one testing session. Training and Game loads can be measured using an internal measure approach with tools like heart rate monitors (Polar, FirstBeat) or an external approach using LPS or GPS systems (e.g., Kinexon, Catapult). As sport science and technology evolves, so does the complexity. This may make analysis a difficult task for a performance health staff. For example, portable force plates may be used to track performance changes over the season such as fatigue, power production, etc. Despite this increasing complexity, Sparta Science, a sport technology company with an emphasis in force plate and jump monitoring uses the simple approach of 'Load, Explode, and Drive'. These three variables are meant to encompass the various phases of a countermovement jump in a simplified manner. Load is a measure of how much force and how quickly it is created. Explode is a measure of how much force and its efficiently in transferred. Drive is a measure of how much force is produced and how long that force is applied. This can be contrasted to the approaches of other companies that have evolved in being able to quantify each granular phase of a jump such as eccentric duration, peak force vs average force, or asymmetries in the eccentric vs concentric phases. Similarly, LPS systems have also evolved from tracking basic metrics such as total distance ran and maximal speed attained, to more complex systems. These complex systems are able to use accelerometers and gyroscopes to track changes in direction, changes in accelerations, and jumps through training sessions and games. As complexity evolves, so too do the methods and tools sports scientists have at their disposal to analyze the data and provide meaningful insight. Throughout the literature review of this project, it became clearer that variables investigated within the field have been derived from measures of correlations and previous research. For example, studies used past research to justify their variables of importance. While this is an acceptable practice, correlations between two variables such as total distance and total workload are starting points for further research rather than standard operating procedures. The current literature centers around using the common theme of using variables of importance set forth by practitioner beliefs or relying on previous investigations. However, in these previous investigations there doesn't seem to be objective reasoning from a mathematical standpoint as to how these variables were chose. There is scarcity in being able to describe relationships between Kinexon variables and how each variable can be used to create unrelated principal components to begin sorting through their importance. However, the studies within the field of external workload share the common theme being based on previous research of lower statistical power. Thus, the implications of the current study are vast because it adds to the field of sport science and external workload management by setting a baseline of important variables within professional basketball. This study adds to the field of basketball performance a set of specific variables; Accumulated Acceleration Load, Mechanical Load, Distance (mi.), Decels Load, Accelerations, Met Power Time High (Sec.), Time (Sec.), Decelerations, Speed (max.) (mph), Speed (% of max.), High Metabolic Power Distance (mi.), Sprints, Metabolic Work (kcal), Acceleration (max.) (ft/s^2), Metabolic Power (max.) (W) that are a combination of previously used ones with newly discovered variables. Furthermore, the use of LPS units are becoming more commonplace within the collegiate level of competition and have already been established at the professional level. Thus, giving performance health staffs a set of variables that are commonplace, accurate, and formed through rigorous statistical methods will be valuable to help optimize on-court performance.

Limitations

This study is not without limitations, which primarily include amount of data, population of interest, and system used. More data is generally agreed to possess more statistical power and accuracy. Although this study began with a robust amount of data points, the majority were discarded due to the high correlation between variables. Even though this resulted in a more accurate model due to low multi-collinearity, it may have compromised the conclusions drawn from the final analysis. The population of professional basketball players was interesting because although assumed to only exist in the United States, the game is global with professional leagues also following globalization. The level of competition differs greatly though, with the highest level of competition and recognition occurring in the NBA (United States). This study was the first of its kind to study external workload variables within NBA players, however the sample size was relatively small due to the logistical challenge of recruiting subjects outside of the author's organization. Lastly, although the PCA proved to be an acceptable and insightful statistical method, the results may have differed if another technique such as a confirmatory factor analysis would have been used.

Recommendations for Future Research

The field of sport science is young and evolving from a technical, tactical, and educational standpoint. This means the opportunity for scholarship and research are also evolving. For example, the amount of research being put out by companies continues to grow because companies now see the value of having evidence backed products and platforms.

Companies will lend their products to academic institutions or sporting organizations for validation and consumer feedback. The rise of 'evidence-based practice' and 'evidence informed practice' has pushed research forward. In the tactical sense, sporting organizations see the importance of being able to quantify game information such as box scores, which contain individual and team information, over the course of the season and against specific opponents. Within NBA practice facilities, it is common to have camera systems over the court to track player movement and shot accuracy. This video is then used by the coaching staff to go over ways of improving player performance through better movement and understanding of game scenarios. As we can see, the technical and tactical components of sports science come back to leveraging data and information to win more games on the court and make championship runs. Similarly, the educational model is following closely behind. At the time of this writing, there are two master's level programs and one doctoral level program in the United States in sports science. The University of Southern California (USC) offers a Master of Science in Biokinesiology with an emphasis in sports science. The University of Pittsburgh has a Master of Science in Sports Science, and Liberty University has a PhD in Health Science with an emphasis in Sport and Exercise Science. At the master's level, curricula focus on research methods, statistics, and learning how to interact with various technologies used within the field such as fore plate testing and player tracking. An internship or clinical rotation is also a foundation of these programs to bridge the gap between understanding how something functions and being able to use technology with athletes within the team setting. At the doctoral level, classes explore a wide range of topics which may include a small amount of didactic sports science but do teach statistics and research methods at an advanced level. It is then up to the interested student to pursue a dissertation topic within the field become an academic expert in the matter.

As the interest in sports science grows, so do the tools and their applications. The PCA in this study was only applied to one NBA organization. It would be interesting to see how the same technique would have fared within another organization or another level of basketball. Also, now that the PCA has been applied successfully to an LPS system, it may be applied to another brand of LPS or another type of player tracking technology. The implications after this research project are vast, but may all benefit the field of sports science and human performance.

REFERENCES

Abdi, H. (n.d.). *Factor Rotations in Factor Analyses*. Retrieved July 27, 2023, from http://www.utdallas.edu/

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/WICS.101

Benson, L. C., Räisänen, A. M., Volkova, V. G., Pasanen, K., & Emery, C. A. (2020). Workload a-WEAR-ness: Monitoring Workload in Team Sports With Wearable Technology. A Scoping Review. *Https://Doi.Org/10.2519/Jospt.2020.9753*, *50*(10), 549–563. https://doi.org/10.2519/JOSPT.2020.9753

Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. *Critical Care*, *7*(6), 451–459. https://doi.org/10.1186/CC2401/FIGURES/13

Bishop, D. (2008). An applied research model for the sport sciences. *Sports Medicine*, *38*(3), 253–263. https://doi.org/10.2165/00007256-200838030-00005/METRICS

Bro, R., methods, A. S.-A., & 2014, undefined. (n.d.). Principal component analysis. *Pubs.Rsc.Org*. Retrieved May 27, 2023, from https://pubs.rsc.org/en/content/articlehtml/2014/ay/c3ay41907j

Bryant, F., & Yarnold, P. (1995). *Principal-components analysis and exploratory and confirmatory factor analysis.* https://psycnet.apa.org/record/1995-97110-004?utm_medium=email&utm_source=transaction

Cangelosi, R., & Goriely, A. (2007). Component retention in principal component analysis with application to cDNA microarray data. *Biology Direct*, *2*(1), 1–21. https://doi.org/10.1186/1745-6150-2-2/TABLES/4

Chappell, J. D., Creighton, R. A., Giuliani, C., Yu, B., & Garrett, W. E. (2007). Kinematics and Electromyography of Landing Preparation in Vertical Stop-Jump. *Https://Doi.Org/10.1177/0363546506294077*, *35*(2), 235–241. https://doi.org/10.1177/0363546506294077

Chatfield, C., & Collins, A. J. (1980). Principal component analysis. *Introduction to Multivariate Analysis*, 57–81. https://doi.org/10.1007/978-1-4899-3184-9_4

Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science*, *1*(3), 98–101. https://doi.org/10.1111/1467-8721.EP10768783

Dascalu, C. G., Dascălu, C. G., & Cozma, C. D. (2017). *The principal components analysis-method to reduce the collinearity in multiple linear regression model; application in medical studies*. https://www.researchgate.net/publication/268265047

Dziuban, C. D., & Shirkey, E. C. (1974a). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*(6), 358–361. https://doi.org/10.1037/H0036316

Dziuban, C. D., & Shirkey, E. C. (1974b). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*(6), 358–361. https://doi.org/10.1037/H0036316

Edwards, T., Spiteri, T., Piggott, B., Bonhotal, J., Haff, G. G., & Joyce, C. (2018). Monitoring and Managing Fatigue in Basketball. *Sports 2018, Vol. 6, Page 19*, *6*(1), 19. https://doi.org/10.3390/SPORTS6010019

Gang Su, X. (2009). Linear regression analysis: Theory and computing. *Linear Regression Analysis: Theory and Computing*, 1–328. https://doi.org/10.1142/6986/SUPPL_FILE/6986_CHAP01.PDF

Gonzalez, A. M., Hoffman, J. R., Rogowski, J. P., Burgos, W., Manalo, E., Weise, K., Fragala, M. S., & Stout, J. R. (2013). Performance changes in NBA basketball players vary in starters vs. nonstarters over a competitive season. *Journal of Strength and Conditioning Research*, *27*(3), 611–615. https://doi.org/10.1519/JSC.0B013E31825DD2D9

Heishman, A., Curtis, M., … E. S.-T. J. of, & 2018, undefined. (2018). Noninvasive assessment of internal and external player load: implications for optimizing athletic performance. *Journals.Lww.Com*. https://doi.org/10.1519/JSC.0000000000002413

Heishman, A., Daub, B., … R. M.-J. of S., & 2020, undefined. (n.d.). Monitoring external training loads and neuromuscular performance for division I basketball players over the preseason. *Ncbi.Nlm.Nih.Gov*. Retrieved May 23, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7039036/

Holland, S. M. (2019). *PRINCIPAL COMPONENTS A N ALYSI S (PCA)*.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Https://Doi.Org/10.1214/Aos/1009210544*, *29*(2), 295–327. https://doi.org/10.1214/AOS/1009210544

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). https://doi.org/10.1098/RSTA.2015.0202

Jolliffe, I. T. (1972). Discarding Variables in a Principal Component Analysis. I: Artificial Data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *21*(2), 160–173. https://doi.org/10.2307/2346488

Kherif, F., Learning, A. L.-M., & 2020, undefined. (n.d.). Principal component analysis. *Elsevier*. Retrieved May 23, 2023, from https://www.sciencedirect.com/science/article/pii/B9780128157398000122

Koshkina, M., Pidaparthy, H., & Elder, J. H. (2021). *Contrastive Learning for Sports Video: Unsupervised Player Classification* (pp. 4528–4536). https://github.com/mkoshkina/teamId.

Laffaye, G., Bardy, B. G., & Durey, A. (2007). Principal component structure and sport-specific differences in the running one-leg vertical jump. *International Journal of Sports Medicine*, *28*(5), 420–425. https://doi.org/10.1055/S-2006-924507

Larsen, R., & Warne, R. T. (2010). Estimating confidence intervals for eigenvalues in exploratory factor analysis. *Behavior Research Methods*, *42*(3), 871–876. https://doi.org/10.3758/BRM.42.3.871/METRICS

Lever, J., Krzywinski, M., & Altman, N. (2017). Points of Significance: Principal component analysis. *Nature Methods*, *14*(7), 641–643. https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=15487091&v=2.1&it=r&id=GALE%7CA497319735&sid=googleScholar&linkaccess=fulltext

Liu, R. X., Kuang, J., Gong, Q., & Hou, X. L. (2003). Principal component regression analysis with spss. *Computer Methods and Programs in Biomedicine*, *71*(2), 141–147. https://doi.org/10.1016/S0169-2607(02)00058-5

Lombard, W., Starling, L., Wewege, L., & Lambert, M. (2020). Changes in countermovement jump performance and subjective readiness-to-train scores following a simulated soccer match. *Https://Doi.Org/10.1080/17461391.2020.1757764*, *21*(5), 647–655. https://doi.org/10.1080/17461391.2020.1757764

McCormack, S., Jones, B., Scantlebury, S., Collins, N., Owen, C., & Till, K. (2021). Using Principal Component Analysis to Compare the Physical Qualities Between Academy and

International Youth Rugby League Players. *International Journal of Sports Physiology and Performance*, *16*(12), 1880–1887. https://doi.org/10.1123/IJSPP.2021-0049

Mizels, J., Erickson, B., & Chalmers, P. (2022). Current State of Data and Analytics Research in Baseball. *Current Reviews in Musculoskeletal Medicine*, *15*(4), 283–290. https://doi.org/10.1007/S12178-022-09763-6/METRICS

Morgan, S. J., and, J. M. M.-A. psychophysiology, & 2017, undefined. (2017). Effect of heart rate variability biofeedback on sport performance, a systematic review. *Springer*, *42*(3), 235–245. https://doi.org/10.1007/s10484-017-9364-2

Motoda, H., and, H. L.-C. of I. (Institute of I., & 2002, undefined. (n.d.). Feature selection, extraction and construction. *Citeseer*. Retrieved May 23, 2023, from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=fb1ce1dd44d1911a611304bf705869112153040a

Panoutsakopoulos, V., … N. P.-J. of S. and H., & 2014, undefined. (n.d.). Sport specificity background affects the principal component structure of vertical squat jump performance of young adult female athletes. *Elsevier*. Retrieved May 23, 2023, from https://www.sciencedirect.com/science/article/pii/S2095254614000027

Principal Component Analysis Guide & Example - Statistics By Jim. (n.d.). Retrieved June 3, 2023, from https://statisticsbyjim.com/basics/principal-component-analysis/

Rana, M., & Mittal, V. (2021). Wearable Sensors for Real-Time Kinematics Analysis in Sports: A Review. *IEEE Sensors Journal*, *21*(2), 1187–1207. https://doi.org/10.1109/JSEN.2020.3019016

Richter, C., O’Reilly, M., & Delahunt, E. (2021). Machine learning in sports science: challenges and opportunities. *Https://Doi.Org/10.1080/14763141.2021.1910334*. https://doi.org/10.1080/14763141.2021.1910334

Robbins, S. M., Renaud, P. J., & Pearsall, D. J. (2021). Principal component analysis identifies differences in ice hockey skating stride between high-and low-calibre players. *Sports biomechanics*, *20*(2), 131-149.

Saucier, D. N., Davarzani, S., Burch V, R. F., Chander, H., Strawderman, L., Freeman, C., Ogden, L., Petway, A., Duvall, A., Crane, C., & Piroli, A. (2021). External Load and Muscle Activation Monitoring of NCAA Division I Basketball Team Using Smart Compression Shorts. *Sensors 2021, Vol. 21, Page 5348*, *21*(16), 5348. https://doi.org/10.3390/S21165348

Sedgwick, P. (2013). Convenience sampling. *BMJ*, *347*(oct25 2), f6304–f6304. https://doi.org/10.1136/BMJ.F6304

Seshadri, D. R., Li, R. T., Voos, J. E., Rowbottom, J. R., Alfes, C. M., Zorman, C. A., & Drummond, C. K. (2019). Wearable sensors for monitoring the internal and external workload of the athlete. *Npj Digital Medicine 2019 2:1*, *2*(1), 1–18. https://doi.org/10.1038/s41746-019-0149-2

Shlens, J. (2014). *A Tutorial on Principal Component Analysis*. https://arxiv.org/abs/1404.1100v1

Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39–42. https://doi.org/10.12691/ajams-8-2-1

Shrestha, N. (2021a). Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*, *9*(1), 4–11. https://doi.org/10.12691/ajams-9-1-2

Shrestha, N. (2021b). Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*, *9*(1), 4–11. https://doi.org/10.12691/ajams-9-1-2

Spägele, T., Kistner, A., & Gollhofer, A. (1999). Modelling, simulation and optimisation of a human vertical jump. *Journal of Biomechanics*, *32*(5), 521–530. https://doi.org/10.1016/S0021-9290(98)00145-6

SPSS Statistics Tutorials and Statistical Guides | Laerd Statistics. (n.d.). Retrieved June 9, 2023, from https://statistics.laerd.com/

Statistics, C. R.-S. T. I. J. of, A, S., & 1964, undefined. (n.d.). The use and interpretation of principal component analysis in applied research. *JSTOR*. Retrieved June 9, 2023, from https://www.jstor.org/stable/25049339

Stone, J. D., Merrigan, J. J., Ramadan, J., Brown, R. S., Cheng, G. T., Hornsby, W. G., Smith, H., Galster, S. M., & Hagen, J. A. (2022). Simplifying External Load Data in NCAA Division-I Men's Basketball Competitions: A Principal Component Analysis. *Frontiers in Sports and Active Living*, *4*, 24. https://doi.org/10.3389/FSPOR.2022.795897/BIBTEX

Suhr, D. D. (n.d.). *Principal Component Analysis vs. Exploratory Factor Analysis*.

Svilar, L., Castellano, J., & Jukić, I. (2018). LOAD MONITORING SYSTEM IN TOP-LEVEL BASKETBALL TEAM: RELATIONSHIP BETWEEN EXTERNAL AND INTERNAL TRAINING LOAD. *Kinesiology*, *50*(1), 25–33. https://doi.org/10.26582/k.50.1.4

Svilar, L., Castellano, J., Jukic, I., & Casamichana, D. (2018). Positional Differences in Elite Basketball: Selecting Appropriate Training-Load Measures. *International Journal of Sports Physiology and Performance*, *13*(7), 947–952. https://doi.org/10.1123/IJSPP.2017-0534

Sylta, Ø., Tønnessen, E., … S. S. of sports, & 2014, undefined. (n.d.). From heart-rate data to training quantification: a comparison of 3 methods of training-intensity analysis. *Journals.Humankinetics.Com*. Retrieved May 23, 2023, from https://journals.humankinetics.com/abstract/journals/ijspp/9/1/article-p100.xml

Tharwat, A. (2016). Principal component analysis - a tutorial. *International Journal of Applied Pattern Recognition*, *3*(3), 197. https://doi.org/10.1504/IJAPR.2016.079733

Tobias, S., & Carlson, J. E. (1969). Brief report: Bartlett's test of sphericity and chance findings in factor analysis. *Multivariate Behavioral Research*, *4*(3), 375–377. https://doi.org/10.1207/S15327906MBR0403_8

Tobias, S., & Carlson, J. E. (2010). BRIEF REPORT: BARTLETT'S TEST OF SPHERICITY AND CHANCE FINDINGS IN FACTOR ANALYSIS. *Http://Dx.Doi.Org/10.1207/S15327906mbr0403_8*, *4*(3), 375–377. https://doi.org/10.1207/S15327906MBR0403_8

uk, M. R.-U. http://people. maths. ox. ac., & 2009, undefined. (2009). Principal component analysis. *Aurora.Troja.Mff.Cuni.Cz*. http://aurora.troja.mff.cuni.cz/nemec/idl/09bonus/pca.pdf

U.S. Americans who played basketball 2021 | Statista. (n.d.). Retrieved July 28, 2023, from https://www.statista.com/statistics/191632/participants-in-basketballl-in-the-us-since-2006/

Wang, C., Vargas, J. T., Stokes, T., Steele, R., & Shrier, I. (2020). Analyzing Activity and Injury: Lessons Learned from the Acute:Chronic Workload Ratio. *Sports Medicine*, *50*(7), 1243–1254. https://doi.org/10.1007/S40279-020-01280-1/METRICS

Watanabe, N. M., Shapiro, S., & Drayer, J. (2021). Big Data and Analytics in Sport Management. *Journal of Sport Management*, *35*(3), 197–202. https://doi.org/10.1123/JSM.2021-0067

Weather, I. J.-, & 1990, undefined. (1990). Principal component analysis: a beginner's guide—I. Introduction and application. *Researchgate.Net*. https://doi.org/10.1002/j.1477-8696.1990.tb05558.x

Wu, P. P. Y., Sterkenburg, N., Everett, K., Chapman, D. W., White, N., & Mengersen, K. (2019). Predicting fatigue using countermovement jump force-time signatures: PCA can distinguish neuromuscular versus metabolic fatigue. *PLOS ONE*, *14*(7), e0219295. https://doi.org/10.1371/JOURNAL.PONE.0219295

APPENDIX

Appendix A

Correlation Matrix

		Distance mi	Speed of max	Speed max.mph	Acceleration max.fts²	Accumulated Acceleration Load
Correlation	Distance (mi)	1.000	.416	.417	.389	.999
	Speed % of max	.416	1.000	.999	.741	.432
	Speed max. (mph)	.417	.999	1.000	.741	.433
	Acceleration max. fts²	.389	.741	.741	1.000	.406
	Accumulated Acceleration Load	.999	.432	.433	.406	1.000
	Decels Load	.995	.391	.392	.367	.994
	High Metabolic Power Distance (mi)	.915	.490	.491	.460	.927
	Metabolic Power max. (W)	.305	.216	.217	.216	.310
	Metabolic Work (kcal)	.073	.202	.204	.223	.087
	Mechanical Load	.995	.393	.395	.370	.994
	Accelerations	.969	.452	.453	.434	.974
	Changes of Direction	.680	.435	.437	.451	.690
	Decelerations	.955	.411	.412	.390	.961
	Sprints	.887	.462	.463	.435	.898
	Time (Sec)	.988	.356	.358	.330	.983
	Met Power Time High (Sec)	.970	.363	.365	.331	.970
Sig. (1-tailed)	Distance (mi)		<.001	<.001	<.001	.000
	Speed % of max	.000		.000	.000	.000
	Speed max. (mph)	.000	.000		.000	.000
	Acceleration max. fts²	.000	.000	.000		.000
	Accumulated Acceleration Load	.000	.000	.000	.000	
	Decels Load	.000	.000	.000	.000	.000
	High Metabolic Power Distance (mi)	.000	.000	.000	.000	.000

	Metabolic Power max.(W)	.000	.000	.000	.000	.000
	Metabolic Work (kcal)	.018	.000	.000	.000	.006
	Mechanical Load	.000	.000	.000	.000	.000
	Accelerations	.000	.000	.000	.000	.000
	Changes of Direction	.000	.000	.000	.000	.000
	Decelerations	.000	.000	.000	.000	.000
	Sprints	.000	.000	.000	.000	.000
	Time (Sec)	.000	.000	.000	.000	.000
	Met Power Time High (Sec)	.000	.000	.000	.000	.000

Correlation Matrix

		Decels Load	High Metabolic Power Distance mi	Metabolic Power max.W	Metabolic Work kcal	Mechanical Load
Correlation	Distance (mi)	.995	.915	.305	.073	.995
	Speed % of max	.391	.490	.216	.202	.393
	Speed max. (mph)	.392	.491	.217	.204	.395
	Acceleration max. fts²	.367	.460	.216	.223	.370
	Accumulated Acceleration Load	.994	.927	.310	.087	.994
	Decels Load	1.000	.905	.313	.058	1.000
	High Metabolic Power Distance (mi)	.905	1.000	.336	.189	.907
	Metabolic Power max.(W)	.313	.336	1.000	.473	.312
	Metabolic Work (kcal)	.058	.189	.473	1.000	.058
	Mechanical Load	1.000	.907	.312	.058	1.000
	Accelerations	.966	.954	.373	.159	.967
	Changes of Direction	.659	.712	.329	.266	.660
	Decelerations	.960	.939	.329	.142	.961
	Sprints	.875	.982	.306	.165	.876
	Time (Sec)	.993	.863	.274	.006	.993
	Met Power Time High (Sec)	.968	.928	.349	.106	.969
Sig. (1-tailed)	Distance (mi)	.000	.000	<.001	.018	.000

	Speed % of max	.000	.000	.000	.000	.000
	Speed max. (mph)	.000	.000	.000	.000	.000
	Acceleration max. fts²	.000	.000	.000	.000	.000
	Accumulated Acceleration Load	.000	.000	.000	.006	.000
	Decels Load		.000	.000	.047	.000
	High Metabolic Power Distance (mi)	.000		.000	.000	.000
	Metabolic Power max.(W)	.000	.000		.000	.000
	Metabolic Work (kcal)	.047	.000	.000		.046
	Mechanical Load	.000	.000	.000	.046	
	Accelerations	.000	.000	.000	.000	.000
	Changes of Direction	.000	.000	.000	.000	.000
	Decelerations	.000	.000	.000	.000	.000
	Sprints	.000	.000	.000	.000	.000
	Time (Sec)	.000	.000	.000	.432	.000
	Met Power Time High (Sec)	.000	.000	.000	.001	.000

Correlation Matrix

		Accelerations	Changes of Direction	Decelerations	Sprints	Time Sec
Correlation	Distance (mi)	.969	.680	.955	.887	.988
	Speed % of max	.452	.435	.411	.462	.356
	Speed max (mph)	.453	.437	.412	.463	.358
	Acceleration max fts²	.434	.451	.390	.435	.330
	Accumulated Acceleration Load	.974	.690	.961	.898	.983
	Decels Load	.966	.659	.960	.875	.993
	High Metabolic Power Distance (mi)	.954	.712	.939	.982	.863
	Metabolic Power max. (W)	.373	.329	.329	.306	.274
	Metabolic Work (kcal)	.159	.266	.142	.165	.006
	Mechanical Load	.967	.660	.961	.876	.993
	Accelerations	1.000	.710	.965	.922	.937
	Changes of Direction	.710	1.000	.680	.691	.617
	Decelerations	.965	.680	1.000	.913	.934
	Sprints	.922	.691	.913	1.000	.832
	Time (Sec)	.937	.617	.934	.832	1.000
	Met Power Time High (Sec)	.963	.676	.960	.906	.950
Sig. (1-tailed)	Distance (mi)	.000	<.001	.000	<.001	.000
	Speed % of max	.000	.000	.000	.000	.000
	Speed max. (mph)	.000	.000	.000	.000	.000
	Acceleration max. fts²	.000	.000	.000	.000	.000
	Accumulated Acceleration Load	.000	.000	.000	.000	.000
	Decels Load	.000	.000	.000	.000	.000
	High Metabolic Power Distance (mi)	.000	.000	.000	.000	.000
	Metabolic Power max. (W)	.000	.000	.000	.000	.000
	Metabolic Work (kcal)	.000	.000	.000	.000	.432
	Mechanical Load	.000	.000	.000	.000	.000
	Accelerations		.000	.000	.000	.000
	Changes of Direction	.000		.000	.000	.000

	Decelerations	.000	.000		.000	.000
	Sprints	.000	.000	.000		.000
	Time (Sec)	.000	.000	.000	.000	
	Met Power Time High (Sec)	.000	.000	.000	.000	.000

Correlation Matrix

		Met Power Time High Sec
Correlation	Distance (mi)	.970
	Speed % of max	.363
	Speed max. (mph)	.365
	Acceleration max. fts²	.331
	Accumulated Acceleration Load	.970
	Decels Load	.968
	High Metabolic Power Distance (mi)	.928
	Metabolic Power max. (W)	.349
	Metabolic Work (kcal)	.106
	Mechanical Load	.969
	Accelerations	.963
	Changes of Direction	.676
	Decelerations	.960
	Sprints	.906
	Time (Sec)	.950
	Met Power Time High (Sec)	1.000
Sig. (1-tailed)	Distance (mi)	.000
	Speed % of max	.000
	Speed max. (mph)	.000
	Acceleration max. fts²	.000
	Accumulated Acceleration Load	.000
	Decels Load	.000
	High Metabolic Power Distance (mi)	.000
	Metabolic Power max. (W)	.000
	Metabolic Work (kcal)	.001
	Mechanical Load	.000
	Accelerations	.000
	Changes of Direction	.000
	Decelerations	.000
	Sprints	.000
	Time (Sec)	.000
	Met Power Time High (Sec)	