

Running Back Performance Predictors: A Cluster Analysis

Andrea Wright

SLM 538

05 May 2023

Note: SLM538 is a class titled “Fundamentals of Coaching.” For this project, I was required to develop a strength and conditioning workout for athletes using some form of analytics.

Abstract

The NFL Combine is hosted each year as a showcase for college athletes hoping to join professional teams. Research has shown correlations between Combine results and career performance, most notably for running backs. This study examined top performers currently playing in the NFL and created four clusters defining the running backs: Dual Threats, Rugged Running Backs, Power Players, and Mid-Range RBs. Then, members of the 2021 draft class were assigned to these clusters and corresponding workouts were written for them to best emphasize their traits as defined by their cluster.

I. Introduction.....	1
II. Literature Review.....	2
III. NFL Combine and Exercise Physiology.....	2
a. Weight.....	2
b. 40-Yard Dash.....	2
c. Cone Drill.....	3
IV. Method.....	3
a. Athlete Selection and Data Methods.....	3
b. Data Collection.....	3
V. Results.....	4
a. K-Modes Clustering.....	4
VI. Strength Programming.....	5
a. Power Players: Chris Evans and Jake Funk.....	5
b. Mid-Range RBs: Khalil Herbert.....	6
VII. Discussion.....	7
a. K-Modes Clustering.....	7
b. Power Players: Chris Evans and Jake Funk.....	7
c. Mid-Range RBs: Khalil Herbert.....	8
VIII. Limitations.....	8
IX. Sources.....	8

I. Introduction

The NFL Combine is an annual, week-long showcase during which athletes undergo a variety of physiological and skill-specific testing in an effort to demonstrate their abilities. It is no question that the professional scouts attending this event have methods of evaluating athletes. However, with the increasing accessibility of analytics, it is possible to enhance these methods. The following research seeks to demonstrate how performance analytics can be implemented from a strength and conditioning standpoint to classify and grade hopeful future NFL running backs.

II. Literature Review

Significant literature examining the efficiency and effectiveness of the NFL Combine exists. A 2016 study examining NFL Combine performance on future performance of running backs and wide receivers suggests that for running backs, the 10-yard dash, a measure of anaerobic capacity, was the most significant predictor. For wide receivers, vertical jump, a measure of muscular power and a skill specific to this position, was the most significant (Teramoto et al., 2016). Further, Combine measures were used to predict All-Pro or Pro Bowl nominations. These results suggest that both physiological measures (height, weight) and performance metrics (40-yard dash results, vertical jump, etc.) have significant impact on successful career performance (Hedlund, 2018).

A study examining Combine results and NFL Draft round suggests that the most significant tests distinguishing drafted versus undrafted players were the 40-yard dash, vertical-jump, pro-agility shuttle, and cone drill. More specifically, these tests were highly successful in determining draft round predictions for running backs, wide receivers, and defensive backs. One limitation of these drills, however, is that they are tests of anaerobic capacity, power, and agility - all relevant to these positions. It is not a surprise that these tests, when applied to offensive and defensive linemen who value strength over speed and power, were not significant in predicting the draft round.

Of all studies examining Combine measurements and performance, running backs consistently appear. A study from *Harvard Sports Analysis* confirms previous literature suggesting that the 40-yard dash, a measure of anaerobic capacity, is the most important, followed by weight, and the Cone Drill. This study suggests that the Combine does not have significant predictive value for wide receiver performance (*See Section VIII. Limitations for further discussion*).

III. NFL Combine and Exercise Physiology

The NFL Combine merges facets of traditional exercise testing with football-specific drills. For running backs, weight, the 40-yard dash, and the cone drill have shown the greatest relationship to career performance.

a. Weight

Weight alone can be a misleading measure when evaluating athletic performance. An athlete with more or less weight does not necessarily equate to more or less athletic ability. A more accurate measure would be body composition, which evaluates the distribution of muscle mass, bone mass, fat mass, and, in some instances, body water. However, this is not to say weight should be eliminated as an element of the NFL Combine. In combination with strength and power drills, such as the bench press or 40-yard dash, distributions of muscle versus fat mass can be surmised. In a strength and conditioning setting after the NFL Draft, the athlete would likely undergo bioelectrical impedance analysis or body plethysmography to more accurately assess body composition.

b. 40-Yard Dash

The 40-yard dash is a 40 yard, straight-line sprint. Sprint evaluations such as this one can be found in traditional exercise testing batteries as a measure of anaerobic power. Anaerobic power, or the work generated by non-oxygen utilizing energy creation systems over some unit time, is a valuable component of a running back's skill set. Athletes that must generate a great amount of force in little time, such as sprinters or powerlifters, are said to possess high anaerobic power. A running back would fall into this category of athlete. Contrast this with a marathon runner, who likely has very little anaerobic power.

c. Cone Drill

The cone drill in the NFL Combine is a measure of agility and quickness. This drill requires an athlete to maneuver quickly around cones, switching direction at high speeds. This is an important skill to a running back who may seek to "trick" their opponent. Unlike the 40-yard dash, this test, despite requiring high speeds by the athlete, *is not* a measure of power. The cone drill is a combination of speed, agility, and body control, all important position-specific skills for a running back.

IV. Method

a. Athlete Selection and Data Methods

In order to examine the college athletes, “top performers” from their respective positions were selected. Please note that in this study, running backs are the position of interest (*See VIII. Limitations for further discussion*). In this case, “top performers” are the top 32 running backs with the most yards in a season, for the 2001-2020 seasons. This number of performers and time frame was selected in order to generate enough observations (performers) to successfully utilize the data evaluation method. There are a select few quarterbacks who consistently generate large numbers of rushing yards that rival athletes whose position title is running back. For example, Baltimore Ravens quarterback Lamar Jackson generated enough rushing yards in both the 2019 and 2020 seasons that he was ranked 6th and 9th, respectively, in rushing yards. Quarterbacks who follow Jackson’s example are eliminated from the dataset.

It was also necessary to account for repeat top performers. For example, Marshawn Lynch is a top running back in 2010, 2011, and 2012. However, he is considered a single data point in this data because he has one set of Combine results. After accounting for quarterbacks and repeat performers, 96 elite, unique running backs were selected from the 2010 to 2020 seasons.

The NFL Combine does not require that all athletes complete all tasks. This reduces the size of top performer reference data - not all running backs completed the 40-yard dash, weight measurement, and cone drill. There were 76 final observations (top-performing athletes) used in cluster analysis.

The top performers were clustered using K-Means clustering. This method of clustering evaluates each observation, or performer, with the goal of distinguishing them into k- groups that have a significant enough difference between them such that they can be considered distinct from one another. Established clusters were saved and named based on these distinguishing factors.

Newly drafted athletes were considered from the 2021 draft class. First, running backs that completed the three tasks of interest (40-yard dash, weight, and cone drill) were subset from the entire draft class. These athletes were then selected at random and evaluated to determine into which cluster they best fit.

b. Data Collection

Combine data was collected using the python library *nfl-data-py 0.3.0*. This library pulls from the popular R database, *nflfastR*, which collects play-by-play data, combine data, and biographical data from the NFL for public use. More information on *nfl-data-py 0.3.0* can be found [at this link](#).

Top performers were collected from Pro-Football-Reference, a public reference database of NFL statistics. This collection was done manually, selecting the top 32 players based on rushing yards from 2001 to 2020.

V. Results

a. K-Modes Clustering

Clustering generated three distinct groups within the top performing running backs. These are displayed in Table 1.

Table 1. Top performer cluster results

Cluster Number	Number of RBs	Cluster Name	40-yd Dash Time [s]	Weight [lbs]	Cone Drill Time [s]
0	16	Dual Threat	4.52	202.4	6.90
1	35	Rugged RBs	4.52	216.5	7.01
2	3	Power Players	4.60	251.7	7.19
3	22	Mid-Range RBs	4.57	228.4	7.12

Clustering resulted in four distinct groups of running backs. Cluster 0 has been called the “Dual Threat.” These athletes are the lightest of all clusters and possess speed and power (as evidenced by a low 40-yard dash time) and impressive agility.

Next, Cluster 1 refers to a group called the “Rugged RBs.” These athletes, like those in Cluster 0, have impressive 40-yard dash and cone drill times. They are slightly heavier than Cluster 0. Cluster 2 is the smallest cluster, with only 3 “top performing running backs” falling into this category. They are the heaviest of all clusters and possess decent 40-yard dash speeds. They have the worst agility times.

Lastly, Cluster 3 is called the Mid-Range Running Backs. They rank in the middle for 40-yard Dash time, weight, and cone drill time. After cluster creation, the draft picks from the 2021 class were assigned to their clusters using the KMeans algorithm. Results can be found in Table 2.

Table 2. Draft pick cluster assignments

Athlete Name	40-yd Dash Time [s]	Weight [lbs]	Cone Drill Time [s]	Cluster Assignment
Chris Evans	4.50	211.0	6.83	1
Jake Funk	4.49	204.0	6.71	0
Khalil Herbert	4.46	210.0	6.96	1
Rhamondre Stevenson	4.64	231.0	7.09	3
Michael Carter	4.50	201.0	6.81	0

Two athletes, Chris Evans and Khalil Herbert, were assigned to Cluster 1, Rugged RBs. This would suggest that these athletes are quick and agile, but heavier than those in Cluster 0. Two athletes, Jake Funk and Michael Carter were placed in Cluster 0. The last athlete, Rhamondre Stevenson, was placed in Cluster 3, Mid-Range Running Backs. Results are evaluated in greater detail in *Section VII. Discussion*.

VI. Strength Programming

a. (Dual Threat) Cluster 0 and (Rugged RBs) Cluster 1: Chris Evans, Khalil Herbert, Jake Funk, and Michael Carter

These athletes were assigned to the cluster that emphasizes anaerobic power. The following workout is an example of what these athletes might do to focus on this aspect of performance. It is important to note that although a running back may be focused on large, powerful movements, they should not completely neglect agility drills, which are still crucial to their position. This becomes more relevant during later periods of pre-season training that are closer to season start.

TIME PERIOD: Early pre-season. A solid conditioning base has been established. Position specific skills are beginning to be implemented into practice. Position specific and team practice are not yet occurring.

LENGTH OF PRACTICE: 1.5 hours

SETTING: Indoor or outdoor field/turf.

WHO: Dual Threat and Rugged RBs. Running backs who fall into the cluster emphasizing high anaerobic power.

PRACTICE GOAL: Continue developing power, but do not neglect agility or relevant skills. This type of running back is best suited for powerful plays that require intense speed and the ability to break through any defensive barrier. Skills in the second half of the workout focus on ball security, which is essential - speed and power are not important if an athlete cannot protect the ball!

Table 3. Power Players example workout

TIME	ACTIVITY	DESCRIPTION/PURPOSE	FEEDBACK + DISCUSSION
0:00 - 0:05	Warm-up jog	Easy, light jog around perimeter of practice area for ~5 minutes.	N/A
0:05 - 0:20	Dynamic warm-up	Dynamic warm-up, including whole-body movements, especially thoracic rotation.	After group warm up, encourage athletes to stretch/readdress anything that they feel needs extra attention.
0:20 - 0:30	Med Ball Bench	S&C + MENTAL Athlete lays on the ground, hands ready to catch the med ball. A partner stands behind them, at their head, and drops the ball. Athlete catches ball, elbows to the ground, and throws ball back up. 4 sets, 8 reps. 1 min rest bw sets. RPE* should be 13-15. Increase med ball weight or encourage speed if RPE is not met.	This is a tough exercise, continue to encourage and motivate athletes. This most directly impacts upper body power, but is also a mental exercise in anticipation and coordination.
0:30 - 0:35	BREAK		
0:35 - 0:45	Resisted sprints (with sleds) increasing weight.	S&C + MENTAL Athletes wear a harness and are attached to sled. They sprint at max effort for 30 yards. Rest for 1 minute, repeat until 10 minutes is over. RPE* 15-17. Weight on sled can increase to reach appropriate RPE.	Remind athletes to pump with arms, have a slight forward lean, and to keep their head up. Again, this can be a tough exercise, especially if weight or speed is increased such that RPE exceeds 17. Encouragement is necessary.

			This is, most directly, lower body power exercise, but also a incorporates an athlete's self-efficacy or belief that they can complete this challenging task.
0:40 - 0:50	Resisted sprints (with sleds) for time	<p>S&C + MENTAL Same as above, but significantly reduce weight on sled, likely 50% of the max weight used in prior exercise.</p> <p>In this version, athletes will sprint as fast as they can with the sled (with reduced weight, this may seem easier + athletes will have greater speed).</p> <p>RPE 15-17.</p>	<p>By reducing the weight, we have removed resistance and are now working on speed + power, rather than just power.</p> <p>Motivation is key! Coming out of a power exercise, where there was significant resistance, this may appear to be a "break." However, max effort should still be exerted so speed AND power are practiced.</p>
0:50 - 0:55	BREAK		
0:55 - 1:00	Ball Security - Switching Sides	<p>TECHNICAL Cones are set up ~1 yard apart. Athlete must weave between the cones, switching the side they hold the ball with each cone. Athlete begins with the ball one on side of the body.</p>	Also a test of coordination, this drill is practice in maintaining control of the ball while moving with great power and speed.
1:00 - 1:10	Ball Security Drill - With a partner	<p>TECHNICAL Athlete moves at ~25% speed, holding the ball. A partner moves alongside, attempting to punch the ball out of the RB's hand. Repeat for both sides.</p> <p>If the partner can effectively keep up, speed can be increased.</p>	This more closely mimics gameplay, as a defensive player may try to punch the ball out of the player's arms.
1:10 - 1:25	Ball Security Drill - switching sides, two partners	<p>TECHNICAL A combination of the last 2 drills, 2 partners line up on either side of the cones. RB weaves through, as with the first version and on either side, partners hold blocking pads and jostle RB as they run through the drill.</p>	<p>Although this drill does not include the punching from the second drill, it does mimic the distractions/interference that will occur during a game.</p> <p>Continue to motivate athlete despite the distractions they are facing.</p>

1:25 - end	BREAK / REGROUP
------------	------------------------

* The Borg Rating of Perceived Exertion (RPE) scale is used here.

b. Mid-Range RBs: Rhamondre Stevenson

Stevenson was placed in the cluster that described well-rounded running backs. The following workout is an example of what he may do in the early off-season.

TIME PERIOD: Early pre-season. A solid conditioning base has been established. Position specific skills are beginning to be implemented into practice. Position specific and team practice are not occurring.

LENGTH OF PRACTICE: 1.5 hours

SETTING: Indoor or outdoor field/turf.

WHO: Mid-Range RBs. Running backs who have solid skills in all relevant aspects, but do not possess extreme speed and power or agility.

PRACTICE GOAL: Continue working on power, speed, and agility - all aspects of a well-rounded running back (which is also best describes this cluster of athletes). Because it is early pre-season, continue with conditioning foundations and then incorporate skill-specific drills.

Table 4. Mid-Range RBs example workout

TIME	ACTIVITY	DESCRIPTION/PURPOSE	FEEDBACK + DISCUSSION
0:00 - 0:05	Warm-up jog	Easy, light jog around perimeter of practice area for ~5 minutes.	N/A
0:05 - 0:20	Dynamic warm-up/ stretching	Dynamic warm-up, including whole-body movements. May also include light agility work such as ladder drills, cone drills, etc.	N/A
0:20 - 0:30	Ball Slams	S&C Athletes take a weighted med ball and slam into the ground. 4 sets, 10 reps. 1 min rest bw sets.	This drill has a focus on upper body power. A low-moderate RPE is recommended here to not exhaust the athlete, but appropriately challenge them and continue building power.

		RPE 14-17. Increase med ball weight if necessary, but be mindful of any danger that may occur by lifting a weighted ball over the head.	
0:35 - 0:40	BREAK		
0:40 - 0:50	Sled Push / Pull	<p>S&C Athlete push a sled 30 yards, rest for 1 minute at the end of each push. Then, pull the sled (moving backwards) for the same 30 yards. Rest 1 minute.</p> <p>RPE 13-16. This RPE is somewhat difficult, but should not exhaust the athlete.</p>	This drill has a focus on lower body power. Again, a low RPE is suggested to not exhaust the athlete but continue to challenge them.
0:50 - 0:55	BREAK / REGROUP		Hydration, praise for hard work in the S&C phase.
0:55 - 1:00	Read and Cut - Direction only, no ball	<p>MENTAL / TECHNICAL Athlete sprints in a straight line towards coach or partner. Coach indicates right or left, where the athlete should cut and run.</p>	Simple start to the technical portion of the workout. Athlete will focus both on reaction time as they approach their partner and technical ability as they successfully cut away.
1:00 - 1:10	Read and Cut - Direction only, with ball	<p>TECHNICAL Same as above, but athlete must carry a ball. Athlete should also be sure to carry the ball on their outside arm, depending on the direction they are indicated by coach.</p>	Progression by adding the ball more closely mimics actual gameplay. In this phase, the athlete must successfully cut away, but also remember to switch their ball-carrying arm.
1:10 - 1:25	Read and Cut - Defender, with ball	<p>TECHNICAL Same as above, but coach is replaced by a defender. Defender will attempt to block RB from running forward, and follow them. Best mimic of actual gameplay.</p>	Best example that mimics gameplay. Athlete must incorporate all of the above aspects, but also compete against a defensive player who is attempting to swipe the ball.
1:25 - end	BREAK AND REGROUP		Emphasize strengths from technical portion of workout and overall good work.

VII. Discussion

a. K-Means Clustering

Cluster 0 and Cluster 1, Dual Threat and Rugged RBs, possess the combination of and quick 40-yard dash times. This suggests that a significant portion of their weight can be attributed to muscle mass as opposed to fat or bone mass. Additionally, it can be surmised that they are highly anaerobic athletes - their speed and power is derived from large stores of intramuscular Phosphocreatine content.

Cluster 3, Mid-Range Running Backs, do not show any extreme qualities in these Combine tests. However, this type of running back is still valuable in the professional football setting - the clusters were determined on “top performers” in the NFL, so it is understood that this type of running back both exists in the NFL, and is successful. It is possible that running backs that fall into this cluster are advantageous in play-action plays, where the quarterback pretends to throw the ball, but has actually handed it off to a running back. In this situation, a certain degree of acting skills or an understanding of the play (and where an optimal space to run might exist) would surpass the skills tested by the Combine.

VIII. Limitations

Limitations of this study range from the technical to the theoretical. Clustering techniques used here are adequate, however, the number of observations of top performers is limited. Clustering works best with a large number of observations. Further studies could expand the range of top performer selection beyond 2010 *or* utilize other Combine drills to evaluate running backs.

From a theoretical standpoint, it should be noted that performance is multi-faceted. An athlete's performance in the NFL Combine occurs almost nearly in a vacuum. The Combine is held indoors at a stable temperature, after weeks of preparation by the athlete, with no defensive players, and relatively little crowd noise. This does not mimic gameday conditions and as such, Combine measurements should not be considered a direct reflection of game performance, rather an indicator of the potential of an athlete.

There is a certain degree of social and psychological consideration that should be taken into account before determining whether an athlete can be successful at a professional level. On the player level, self-regulatory efficacy of an athlete can be a determining factor in successful performance. Additionally, competitive efficacy, especially for a running back, could influence their success. For example, Chris Evans was defined as a Rugged RB, implying that his position as a running back would

likely be in a situation that required him to literally break through his opponents as he carries the ball. If Evans, for any reason, felt hesitation or doubted his ability to compete, he may not break through defensive players. No amount of physical strength and conditioning workouts, no matter how much they emphasized strength and power, could defeat this fear.

American football is a team sport and thus, communication is a powerful tool that can be a determining factor in a team's performance. The offense as a group should have a considerable amount of team cohesion in order to operate successfully. Between running backs and quarterbacks, this relationship is short, lasting barely a few seconds as the quarterback hands off the ball. However, for a running back to trust that the offensive lineman will create a route for him speaks to the respect and trust that must exist for a play to be successful. Team efficacy and cohesion may be a reason why, in many studies, relationships between Combine results and career performance of wide receivers is minimal. As a wide receiver runs far from the quarterback, there must be some form of communication or role clarity that exists such that the wide receiver can catch the ball. The wide receiver must understand that it is his role to get to some predetermined spot. The quarterback must understand that it is his role to get the ball to that spot. Because the relationship appears to have deep roots in the psychosocial realm, it may explain why the Combine is not as successful at predicting wide receiver career performance.

IX. Sources

- Hedlund D. P. (2018). Performance of Future Elite Players at the National Football League Scouting Combine. *Journal of strength and conditioning research*, 32(11), 3112–3118.
<https://doi.org/10.1519/JSC.0000000000002252>
- Lotter, B. (2015, February 23). The NFL Combine Actually Matters. *Harvard Sports Analysis*.
<https://harvardsportsanalysis.org/2015/02/the-nfl-combine-actually-matters/>
- Teramoto, M., Cross, C. L., & Willick, S. E. (2016). Predictive Value of National Football League Scouting Combine on Future Performance of Running Backs and Wide Receivers. *Journal of strength and conditioning research*, 30(5), 1379–1390.
<https://doi.org/10.1519/JSC.0000000000001202>