



# Projecting NFL Potential from College Career Performance Curve

Brian Lehman  
Visionist, Inc.  
brian.lehman@visionistinc.com

## Abstract

Everyone attempts to determine which college football players will succeed at the next level. This paper takes a novel approach to projecting true NFL potential. The projection methods themselves are ordinary, but our foundation is unique. We adapted the proven Elo rating system to develop player ratings that evaluate game performance while also accounting for strength of opponent. Over the last ten draft years, our player Elo ratings alone identified players whose performance value in the NFL was roughly equivalent to those drafted. We use our player performance curves as the basis for projecting NFL potential.

## 1. Introduction

At last year's conference, Mike Leach offered his opinion on the belief that an inaccurate college athlete can be developed into an NFL quarterback, simply stating, "You can't coach accuracy."<sup>1</sup> The selection of a great franchise quarterback can be dramatic. Think about the impact Patrick Mahomes and Lamar Jackson have had on their teams. They help generate more wins, excitement, fans, and revenue. Conversely, a poor first round choice can be catastrophic to an organization. Early picks are costly. If they fail, the time to recover from that bad investment can take years. Even with the high stakes and considerable resources invested to get it right, the NFL draft is still a gamble. Skill levels vary widely across college football and the sample size to evaluate talent is small. This makes predicting NFL future performance difficult.

Each NFL team devotes considerable resources to ensuring draft success. General managers, coaches, and personnel departments pour over scouting reports, combine numbers, pro day data, and video to assess their priority picks. Though the inputs may be the same, team approaches to identifying talent are unique. More and more, organizations are incorporating analytics into their player evaluations in hopes to improve their success rate.

In this paper, we introduce a player rating system that helps level the skill-diverse college football landscape. These normalized player performance metrics enable quantitative player comparisons previously unavailable. Our approach evaluates performance at the game level and tracks that performance over the course of a player's college career. A performance curve is generated from these game-by-game metrics, providing a visual representation of a player's career progression. We postulate that players with a similar development experience can serve as a model to project future performance.

## 2. Background



Our methodology focuses on actual game performance rather than physical attributes, combine numbers, pro day data, school, or any other player metrics. Athleticism, size, and speed are important, but it must translate to game performance. Many of the historic draft fails have been great athletes that didn't play to their potential in college. In almost every case, those same players certainly could not reach that potential against NFL talent.

The challenge of evaluating thousands of players across the skill-diverse college football population in a quantifiable way and enables meaningful comparisons is daunting. To solve this, we adapted concepts of the Elo rating system, developed as a method of calculating the relative skill levels of chess players, and applied them to individual players. The core motivation stems from the key tenets of professor Arpad Elo's methodology; game-by-game player ratings adjust proportionally to the level of competition. Winning a game versus an inferior opponent may raise your rating slightly, but overall is insignificant. Conversely, beating a superior opponent will greatly improve your score. Over time, the rating reflects a player's true skill level, accounting for performance weighted by the strength of competition.

The idea of distilling a player's statistics into a binary outcome may seem overly simplistic. When combined with the Elo concept, this method provides a better metric for comparative player analysis. Think about how teams are evaluated. Teams are judged by overall record, strength of schedule, head-to-head competition, comparative outcomes of common opponents, and the eye test. Football experts, such as the college football playoff committee, use these factors to rank teams.<sup>2</sup> Games provide a discrete boundary for teams to make a statement. For instance, we do not look at cumulative team scores (points for vs. points against) or statistics as the primary tool to evaluate teams. But this is the primary way players are evaluated. Look at any college football statistical leaders page. Players are ranked by overall season statistics. All context of game-by-game performance is lost. Our player Elo model rates player statistics after each game and weights each performance by the strength of opponent.

We use a player's Elo rating history, or college career performance curve, as a way to project their future NFL potential. We propose that players with similar Elo performance progression in college should experience similar success (or failure) at the next level. Our Elo ratings distill various complex and independent factors into a single metric. When viewed as a time series it tells a story. A player may experience a slow start to their college career, ratings-wise, because they are backing up a star and play sparingly or in cleanup action when the game outcome is apparent. In future years, that same player becomes the star. While he waited his turn, maybe he worked hard, developed physically and mentally, then exploded on the scene when he got his chance. The stories may vary, but those with similar performance progression curves share that history.

### 3. Methodology

The following sections will fully describe our methodology: Section 3.1 defines how we generate player Elo ratings; Section 3.2 illustrates the value of the college career performance curve; Section 3.3 details our curve similarity method; and Section 3.4 describes our projection framework.

#### 3.1. Player Elo Ratings

Elo ratings are historically proven to work, so we kept the core algorithm. The significant innovation we pose is translating team competition and outcome space to individual player space.



We accomplish this in a way that maintains the simplicity of the Elo formula and makes good football sense.

### 3.1.1. Standard Elo Rating System

The Elo system was designed to measure the skill level of competitors in zero-sum games. As such, a team's Elo rating increases or decreases based on the outcome against another rated team. After each game, the winner "takes" points from the loser. The difference between ratings determines the number of points gained or lost after each outcome. In the Elo system, performance is inferred from cumulative wins and losses against other teams. Team ratings depend on the ratings of their opponents and the game outcomes. The difference in ratings provides an estimate for the expected score, or probability of outcome, between them. Given Team A and Team B with Elo ratings  $R_A$  and  $R_B$ , the formula for the expected score of Team A is:

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}} \quad (1)$$

Likewise, the expected score of Team B is:

$$E_B = \frac{1}{1 + 10^{(R_A - R_B)/400}} \quad (2)$$

These expected scores, or probabilities of win, for Team A and Team B are then used in the equation to update the team Elo rating ( $R_A$ ).

$$R'_A = R_A + K(G_A - E_A) \quad (3)$$

Where  $R'_A$  is the updated Elo rating and  $G_A$  is the game outcome;  $G_A = 1$  if Team A wins and  $G_A = 0$  if Team A loses.  $K$ , or  $K$ -factor, represents the maximum rating adjustment possible per game.<sup>3</sup> Each application of Elo tends to use their own  $K$ -factor. Variables like number of games/observations and team maturity (new or well-established) drive the choice of  $K$ -factor value.

The updated Elo ratings are computed after each game. If a team does better than expected over time their rating will rise accordingly. Similarly, more losses than expected will lower their rating. A stable Elo rating over time means that a team is performing at their rating level.

### 3.1.2. Adapting for Player Ratings

Wins and losses are foundational elements of the Elo rating system. However, aside from possibly quarterbacks, game outcomes are not typically attributed to individual players. Players are typically evaluated based on their statistics. Each position group has a set of standard statistics that are used to measure their performance.

Our method for player ratings merges the two concepts. We simply classify raw statistics into a win or loss per game. A draw line, or win/loss threshold, is derived for each evaluated statistic. If a player's performance exceeds the statistical draw line, it is considered a win. Otherwise, it is classified as a loss. The formula for a player's Elo rating based on the statistical win/loss is identical to the team Elo systems presented in Section 3.1.1.

To calculate a player's Elo, we need the current Elo ratings of both the player and his opponent. So, who is the opponent? Consider a running back as the player we are evaluating. The opposing team is too general to be considered the opponent, as the team has defense, offense, and special team



components. Even the entire team defense is broader than we want. Our methodology leverages the Elo rating model and applies it to players pitted against an opponent's ability to oppose that player's position (e.g., running back vs. rush defense, receiver vs. pass defense, etc.). In the case of a running back, for each game, the player is evaluated against the opposing team's rush defense.

It should be noted that the team rush defense is evaluated against the opposing team rush offense for the purposes of determining the team's rush defense Elo. We evaluate every team's rush offense, rush defense, pass offense, and pass defense using the same statistical Elo method. These team components are used as the player opponent depending on the statistic being measured.

Player Elo ratings are computed for each statistic independently. When evaluating for a specific position (e.g., running back), the Elo ratings for the relevant statistics (e.g., rush yards per game, yards per carry) are aggregated into a composite player Elo rating. This composite rating currently evenly weights the individual statistics to generate an overall metric for each position.

The Elo ratings system is designed for zero-sum games, the notion that for every winner there is a loser. In keeping with that concept, we use the median as the basis for our statistical draw lines. We want a fixed draw line for each statistic, so we use historic (2000 – 2019) team game statistics to derive our initial median values. This approach yielded draw lines for a selected set of team and individual offensive oriented ratings. Other offensive and defensive oriented draw lines were developed to establish a comprehensive position-by-position win versus loss assessment

We used team and player game statistics from the 2000-2019 college football seasons from Sports Reference.<sup>4</sup> The data covers the NCAA Division I Football Bowl Subdivision (FBS) teams and players, so our results are missing players from the Football Championship Subdivision (FCS). Many FCS players have been drafted and have had rich NFL careers. We dropped these players from the analysis of our results, since we lacked the input data to evaluate them. We believe with the same team and player statistics at a game summary level our methods would work just as well.

### **3.2. College Career Performance Curve**

While a player's most recent Elo rating indicates his current skill level, our player Elo ratings are generated for every game of each player's college career. Using the historical game-by-game ratings, we graph a player's rating progression over time. Each player starts their college career at the baseline Elo rating, in our case 1300. Their rating is updated after every game that player participates in based on their performance and the strength of their opponent. If they perform as expected, the rating remains constant (curve stays flat). Exceeding expectations moves the rating and curve up. A poor game performance relative to opponent strength drives the rating and curve down. The result is a time series of a player's rating. We indicate season breaks with a dot. The numeric rating is arbitrary. The relative rating is what matters. For this reason, we omit Y-axis labels. The X-axis is games played, right justified.

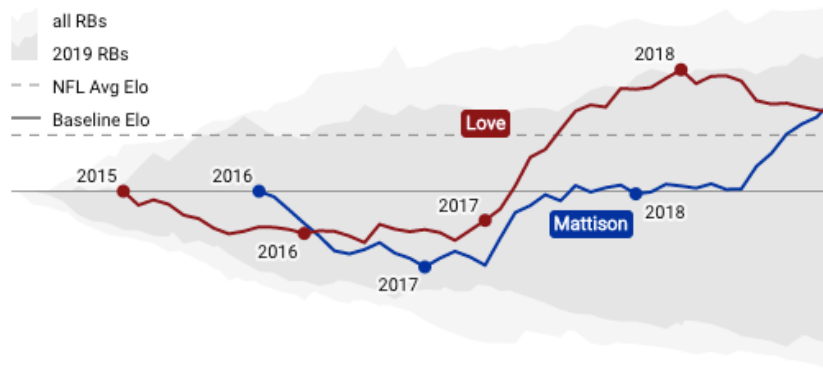


Figure 1 - Performance profiles of Bryce Love (red) and Alexander Mattison (blue)

This college career performance curve offers a unique look at a player’s development. Players with similar final Elo ratings may have taken very different paths to get there. Compare the performance curves of Bryce Love and Alexander Mattison in Figure 1. Love started slow, with his rating settling below the baseline for his first two seasons. His rating skyrocketed his junior season, the year he was the runner up to Baker Mayfield in the Heisman Trophy voting. His rating dipped his final season, as he struggled with injury and teams were ready for him. Mattison also dropped below the baseline his first season. He jumped back to the baseline in his second season, where he remained until halfway through his final season. He finished his college career strong, with an upward trajectory.

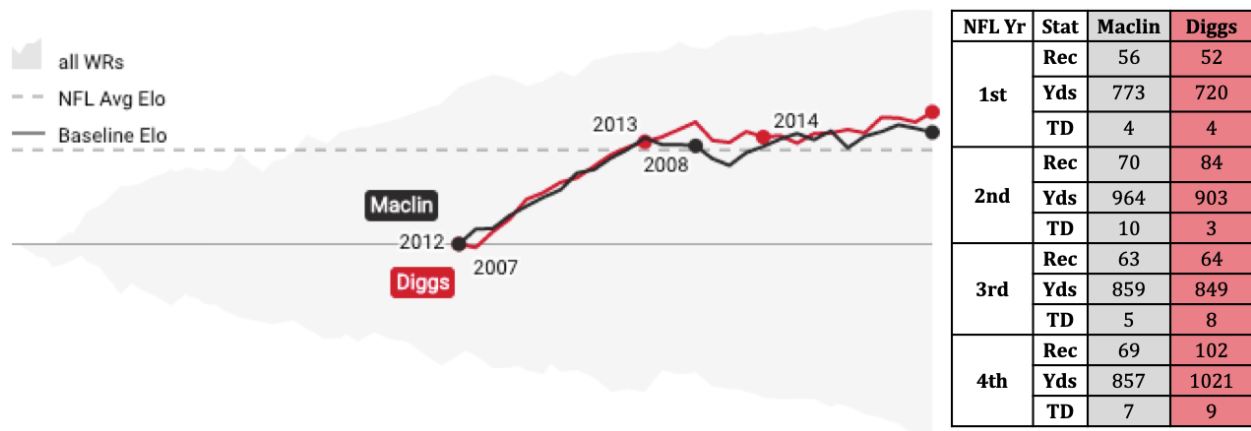


Figure 2 - Stefon Diggs (red) and Jeremy Maclin (black) have nearly identical college career performance curves and equally similar performance in their first four NFL seasons (right)

In other cases, players have similar performance progression throughout their college career. In one example, where the career curves are nearly identical (Figure 2), the NFL performance of the two players is equally similar. Stefon Diggs’ first four NFL season statistics parallel those of Jeremy Maclin to the catch, yard, average, and touchdown. The discovery of this correlation between college performance curve similarity and future NFL performance similarity became the basis for our approach.



We propose that a player's journey to reach his final college Elo rating is a significant factor in projecting NFL potential. The final player rating proves to be a good predictor of performance at the next level, but there are some false positives using that metric alone. We restricted the model of future performance by considering only players that exhibited a similar college career performance progression. This enabled us to generate a distribution of future potential based on those similar players' actual performance in the NFL.

### 3.3. Curve Similarity

To match players with similar college career performance curves we tested a number of curve similarity algorithms, including correlation, minimum least square error, and  $R^2$ . We also considered different approaches to deal with varying curve lengths, based on differences in number of college games played. We found  $R^2$  worked best for our purposes. It is also used by Google Correlate for comparing time series in search query volume.<sup>5</sup>

To account for varied career lengths when comparing players, we had the most success with extending the shorter curve to match the length of the longest curve. Our goal is to match players at their transition from college to pro football. To achieve that, we align the performance curves at the end of their career (see Figure 1). We prepend the shorter curve with the baseline Elo for comparison purposes. This extension approximates a player with an average start to their college career. To find players who had a similar start to an active college athlete, we would have aligned the curves at the start and shortened the curves of the compared players.

To ensure the integrity of our projections, we only consider players from previous years when running our  $R^2$  similarity. This is important for a couple of reasons. From an application standpoint, an NFL team using these projections certainly can't see the future. The results might be interesting, but we want to show the applied utility of our projections. Secondly, the league has changed over time and forward-looking similarities could skew the projections of those from the past. A great example of this is "air raid" quarterbacks. For instance, Graham Harrell (2009) is one of the quarterbacks similar to Patrick Mahomes (2017), but Mahomes cannot be one of the quarterbacks in Harrell's similarity list.

### 3.4. Projection Framework

Using the career performance curve similarity, we created an NFL projection framework. In addition to the curve, we track each player's NFL experience. Did they get a shot in the NFL? If so, how did they perform? We use Approximate Value (AV) as our metric for value in the NFL. AV attempts to quantify a player's value at the end of every season.<sup>6</sup> Many of the players in our data set are still active, so the cumulative AV values vary greatly. We use a player's average AV per year to essentially normalize this across our data.

For each player, we find all similar players that came before them based on their performance curves. We set a threshold of 0.5 to limit our  $R^2$  matching criteria. The resulting group size varies from player to player. We use the size as a confidence metric. The larger the similarity group, the more confident we are in the projection. We also consider the percentage of players in the group that made the NFL in our projection. The similar players with NFL experience are used in a Monte Carlo simulation to forecast NFL potential.

The NFL similar players' names populate a virtual hat. Each name appears a number of times commensurate with their similarity score (eg. 82 times for an  $R^2$  value of 0.82). We perform ten thousand simulated careers by selecting a random name with uniform probability. For each



selection we use a random gaussian probability distribution centered about the selected player's AV/year and a standard deviation based on the similarity score; the higher the similarity, the tighter the distribution. This second randomization provides an added measure of variability based on football sense. We aren't predicting that player will have the exact AV/year. We instead simulate a random AV/year from a normal distribution based on their similarity. The accumulation of these simulated NFL careers is the target player's NFL performance projection distribution of their forecasted AV/year in the NFL as shown below in Figure 3.

The probability a player achieves a certain status (practice squad, backup, starter, franchise player) is reported as the projection. These categories are delineated by vertical lines in Figure 3. The average percentages at each position for NFL potential are also reported in Table 1. An individual player is compared to the average projections to assess their forecasted NFL potential value. The range of AV/year can be large, and this helps capture a player's expected outcome without becoming too skewed by outliers. It also accounts for the number of similar players as well as how similar they are.

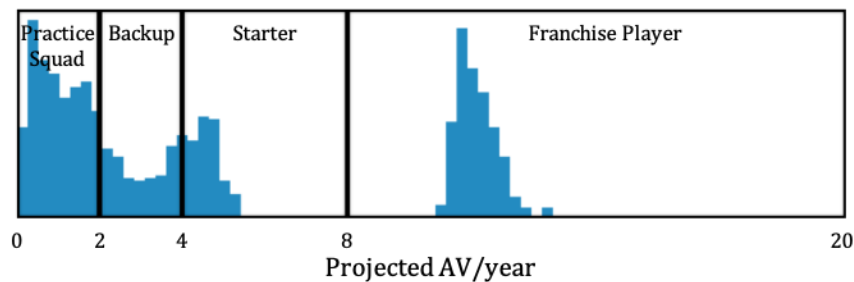


Figure 3 – The projected AV/year distribution for Baker Mayfield based on our simulations

To simplify evaluation of the projection distributions we segmented the distribution into four categories of NFL performance as shown in Table 1. We chose the ranges based on analysis of the AV/year of NFL players and their perceived classification as franchise player, starter, backup, or practice squad player. Table 1 is evidence that distributions tend to be non-normal, as are those of NFL players. This highlights another benefit of using Elo; it can accurately represent any player distribution be it, bi-modal, asymmetric or otherwise non-normal.

Table 1 – NFL performance by position

Category	AV/year	QB	RB	WR	TE
Franchise player	8+	7%	2%	2%	4%
Starter	4-8	17%	9%	10%	6%
Backup	2-4	16%	14%	14%	6%
Practice squad	0-2	60%	75%	74%	84%

## 4. Results

Our player projections build on our novel concept of player Elo ratings. As player Elo is foundational to our player similarity, we will first describe our results of simply ranking players based on their final Elo rating. We then show how the results of our player projection offer greater insight to help teams identify those players with true NFL potential. We focus our discussion on



skill position players because their statistics are more specific to their position. Our method works for all players with box score statistics - everyone but offensive linemen. The system works for defensive players but requires more thorough analysis since defensive statistics are harder to isolate to a specific position.

#### 4.1. Projection from Elo Ratings Alone

Viewed as a game-by-game timeline, the player curves offer many insights into future NFL potential.<sup>7</sup> A great example of the value of our player Elo system can be seen in the 2016 NFL draft. Christian Hackenberg was selected in the second round, but our player Elo ratings would have pointed the Jets toward Dak Prescott instead (Figure 4).

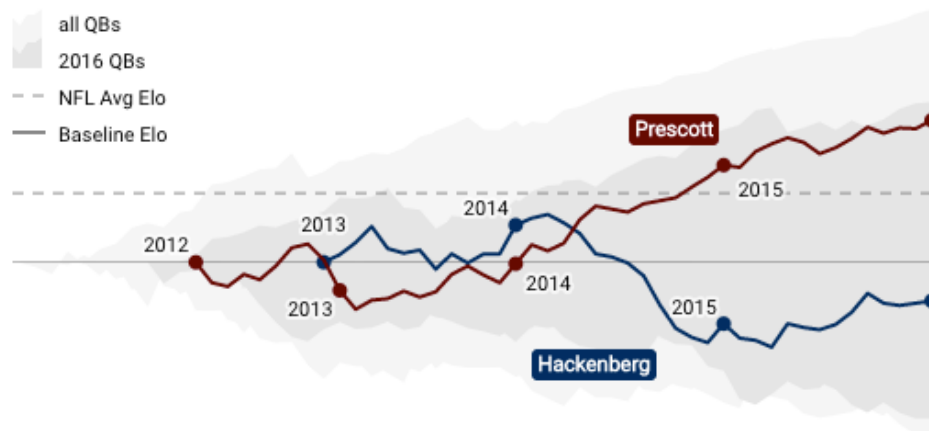


Figure 4 – Comparison of the college career performance curves of Christian Hackenberg (blue) and Dak Prescott (red)

To quantify our results, we compare our top Elo rated skill position players with players drafted in that position. We use AV as the independent measure of NFL performance for the basis of our comparison. For each draft class, from 2008 to 2017, we identified the running backs, wide receivers, and quarterbacks selected. We then compared each year's class with our top Elo rated players from the same year at that position. For instance, in 2015 there were 7 FBS quarterbacks drafted. We, therefore, used our top 7 quarterbacks that ended their college career in the 2014 season for comparison. We compared the cumulative career AV of the drafted players with the cumulative NFL career AV of the same number of our top Elo rated players.

Figure 5 shows the results by skill position by year. Many of the players drafted match our top-rated players, but they are not identical. The graph displays the cumulative AV of the drafted players compared with our top Elo rated players. While the performance of our Elo rating system varies from year to year, the 10-year aggregate performance is impressive. Without spending any time scouting, watching film, or debating talent our results approximate and sometimes exceed the draft results of numerous coaches, GMs, and scouts of the 32 NFL teams. We are not suggesting scouts and GMs be replaced by an algorithm. We are simply attempting to highlight the significance of our results from player Elo alone.



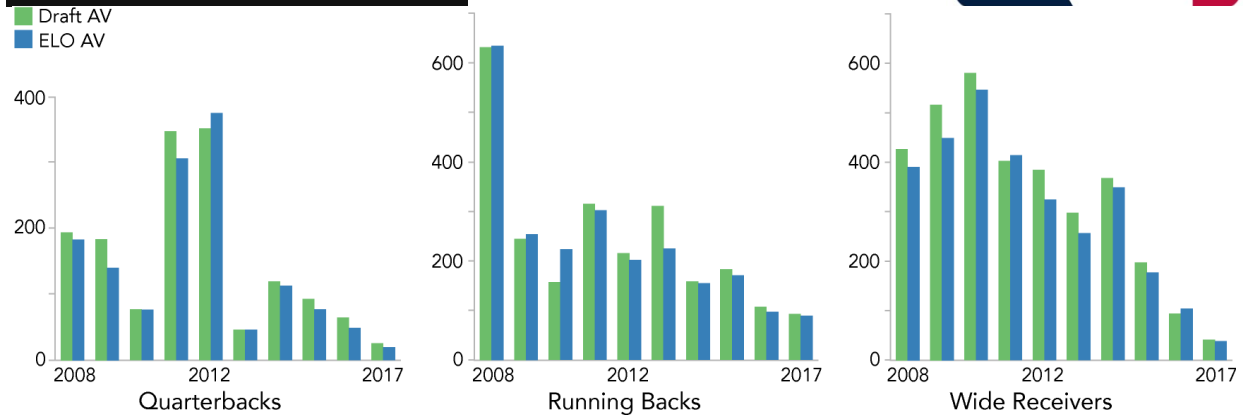


Figure 5 - Comparison of the NFL career Approximate Value (AV) of skill position players drafted with those ranked highest in our Elo ratings from 2008 to 2017.

There are several undrafted players our method has identified that have already or are currently making an impact in the NFL; running backs like Pierre Thomas, LaGarrette Blount, BenJarvus Green-Ellis, and Phillip Lindsay; wide receivers like Danny Amendola, Doug Baldwin, Cole Beasley, and Willie Snead; and quarterbacks like Chase Daniel and Case Keenum. But player Elo alone also identified players that did not have the same level of success, most notably at the quarterback position; Kellen Moore, Colt Brennan, Aaron Murray, Graham Harrell to name a few.

## 4.2. Projection from Performance Curve Similarity

We recognized there was more information in the player performance curves than the final rating that could potentially improve our forecast of NFL talent. Modeling a player’s future performance from those with similar college career development provides some interesting insights. The player projection results are more easily conveyed at the micro versus macro level. We offer specific examples to detail our findings.

### 4.2.1. Quarterbacks (QBs)

Our performance curve similarity exposes an immediate trend with our top Elo rated quarterbacks. Table 2 shows our top fifteen quarterbacks, the number of similar NFL QBs, the total number of similar QBs, and the percentage of NFL QBs in their similarity set. The quarterbacks that did not make it in the NFL have very few players that exhibit a similar college career progression, especially those with NFL experience. Projecting NFL potential from such a small similarity sample size would have low confidence. The final column of Table 2 shows the names of the similar QBs.

Table 2 - The similarity profiles of our top rated quarterbacks based on their performance curves

Quarterback	Similar QBs			Similar NFL QBs <i>*denotes partial list</i>
	NFL	All	NFL/All	
Baker Mayfield	12	15	80.0%	Mariota, Luck, Watson, Harrell, Moore*
Kellen Moore	1	1	100.0%	Tebow
Marcus Mariota	16	20	80.0%	Wilson, Luck, Carr, Bradford, Griffin*
Robert Griffin III	9	13	69.2%	Rodgers, Dalton, Bradford, Smith*
Andrew Luck	8	12	66.7%	Bradford, Daniel, Harrell, Leinart*
Colt Brennan	1	2	50.0%	Leinart
Case Keenum	4	6	66.7%	Leinart, Tebow, Hall, McCoy
Tim Tebow	1	3	33.3%	Leinart



Deshaun Watson	20	24	83.3%	Wilson, Luck, Mariota, Carr, Bradford*
Russell Wilson	11	17	64.7%	Rodgers, Dalton, Bradford, Kaepernick*
Aaron Murray	2	2	100.0%	Keenum, McCoy
Max Hall	5	7	71.4%	Leinart, Harrell, Brohm, Kolb, Daniel
Derek Carr	16	20	80.0%	Wilson, Luck, Foles, Bradford, Dalton*
Graham Harrell	3	5	60.0%	Leinart, Brohm, Kolb

In contrast, we present our projections for a number of NFL starting quarterbacks that are in our sample set. We need to have enough history to be able to compare college career performance curves with previous players, so the earliest draft class we evaluate is 2008. We don't have the necessary data to evaluate some of the older veteran QBs (Brady, Brees, Rodgers, Roethlisberger, Rivers). We also don't have FCS data, so we are missing QBs like Wentz and Garoppolo. Table 3 shows each QB and their projected distribution. The distribution shows simulated AV/year from 0 to 20. The vertical bars represent the bounds of our player classifications that we first presented in Table 1; practice squad, backup, starter, and franchise player.

Table 3 – Our QB projections of NFL potential value using simulations of QBs with similar curves

Quarterback	Similar QBs			Projection				Distribution (AV/year)
	NFL	All	NFL/All	Practice	Backup	Starter	Franchise	
Baker Mayfield	12	15	80.0%	37.7%	12.3%	13.6%	27.7%	
Deshaun Watson	20	24	83.3%	50.5%	13.4%	7.8%	19.2%	
Russell Wilson	11	17	64.7%	39.8%	9.8%	16.3%	21.8%	
Derek Carr	16	20	80.0%	38.0%	15.2%	19.8%	16.7%	
Jared Goff	23	39	59.0%	27.1%	22.3%	23.2%	24.9%	
Sam Darnold	28	51	54.9%	28.9%	19.2%	15.5%	29.7%	
Kyler Murray	14	32	43.8%	22.0%	15.4%	12.0%	38.4%	
Patrick Mahomes	28	50	56.0%	28.1%	20.1%	16.2%	30.9%	
Dak Prescott	30	56	53.6%	35.4%	17.7%	16.8%	20.4%	
Jameis Winston	21	33	63.6%	29.4%	16.5%	25.4%	18.2%	
Mitch Trubisky	12	34	35.3%	36.0%	11.3%	18.0%	24.1%	
Lamar Jackson	25	44	56.8%	24.1%	18.3%	15.1%	35.1%	
Matt Ryan	9	17	52.9%	21.5%	6.3%	27.0%	38.4%	

#### 4.2.2. Running Backs (RBs)



There are many more running backs than quarterbacks, so the similarity set is typically much larger for each player. Statistically, this helps produce a better simulation distribution. Much like QBs, the distribution is rarely, if ever, normal. Even with the increased numbers we can see outliers in our high Elo rated RBs. Table 4 shows our top five RBs based on their final Elo rating.

Table 4 - The similarity profiles of our top rated running backs based on their performance curves

Running Back	Similar RBs			Similar NFL RBs <i>*denotes partial list</i>
	NFL	All	NFL/All	
LaMichael James	4	4	100.0%	McFadden, Rodgers, Hart, Slaton
Toby Gerhart	29	33	87.9%	Drew, Lynch, McCoy, Wells, Pittman*
Derrick Henry	68	81	84.0%	Gurley, Lynch, Bell, Ingram, Hyde, Drew*
Ezekiel Elliott	48	55	87.3%	Gurley, Lynch, Gerhart, Gordon, Rice*
Dalvin Cook	23	26	88.5%	Peterson, Gurley, Lynch, Rice, Sproles*

The relatively low similarity set of LaMichael James is immediately obvious and once again, this correlates to a player whose college success did not translate to the NFL. In Table 5 we show our projections for a number of NFL starting RBs. The AV values for RBs tend to be lower than QBs, which is evidenced in the extent of the RB distributions.

Table 5 - Our RB projections of NFL potential value using simulations of RBs with similar curves

Running Back	Similar RBs			Projection				Distribution (AV/year)
	NFL	All	NFL/All	Practice	Backup	Starter	Franchise	
Derrick Henry	68	81	84.0%	31.8%	23.4%	35.8%	9.1%	
Ezekiel Elliott	48	55	87.3%	29.4%	25.7%	34.8%	10.1%	
Dalvin Cook	23	26	88.5%	21.9%	16.5%	43.2%	18.5%	
Todd Gurley	31	37	83.8%	36.9%	21.1%	38.0%	4.0%	
Leonard Fournette	43	48	89.6%	26.3%	24.7%	34.9%	14.1%	
Melvin Gordon	57	71	80.3%	33.8%	24.2%	33.7%	8.4%	
Mark Ingram	21	27	77.8%	19.6%	19.5%	47.0%	14.0%	
Saquon Barkley	25	35	71.4%	31.6%	23.4%	36.5%	8.5%	
LeVeon Bell	34	50	68.0%	44.2%	22.2%	27.0%	6.6%	
Christian McCaffrey	61	79	77.2%	32.0%	28.9%	30.4%	8.7%	

### 4.2.3. Wide Receivers (WRs)

The trend continues for wide receivers. Many WRs have over one hundred similar players to use for simulation. However, a few have career performance curves that are fairly unique. Just as in the case of QBs and RBs, these WRs tend to be players whose collegiate success did not translate to the



NFL; Jarett Dillard, Ryan Grice-Mullen, Kenny McKinley, Jason Rivers, BJ Cunningham to name a few. One exception to this trend is Davone Bess. Bess only had 11 WRs with similar performance curves, 7 of which with NFL experience. He rated highly in our player Elo ratings and was not drafted. He had a decent NFL career, amassing a cumulative AV of 30 in his six seasons, for an AV/year of 5.

With no Alabama versus Clemson matchup this year, we make a claim for wide receiver university (WRU) by comparing three Alabama NFL WRs with three from Clemson in Table 6. From the distributions, it appears Alabama WRs have a slight edge. Jones and Ridley have fewer similar players to project future performance. All six WRs have already proven themselves in the NFL. Watkins, Ridley, and Williams have AV/year between 6-7, classifying them as “Starters”. Cooper, Hopkins, and Jones have AV/year greater than 8, making them worthy of the “Franchise” tag.

Table 6 – Comparison of our projections for Clemson and Alabama NFL wide receivers

Wide Receiver	Similar WRs			Projection				Distribution (AV/year)
	NFL	All	NFL/All	Practice	Backup	Starter	Franchise	
Mike Williams	160	246	65.0%	42.8%	21.7%	27.5%	8.0%	
Amari Cooper	77	115	67.0%	42.4%	17.2%	30.5%	10.0%	
Deandre Hopkins	82	136	60.3%	46.1%	21.7%	26.7%	5.5%	
Julio Jones	20	36	55.6%	46.3%	15.5%	30.1%	8.1%	
Sammy Watkins	56	87	64.4%	45.9%	17.1%	28.8%	8.2%	
Calvin Ridley	10	21	47.6%	43.6%	20.0%	24.8%	11.5%	

## 5. Discussion

Our player Elo rating system does a good job of evaluating college football players in a way that allows meaningful comparisons. Many NFL teams have noted the power and rarity of this visual comparison. We use those curves to find players with a similar performance profile and model future performance from the players that precede them. Our projections provide additional context that is valuable in forecasting NFL potential. A player with a lower Elo rating may project as a better NFL player. A perfect example is Lamar Jackson. He rates much lower than Baker Mayfield, the top rated QB in our data set. Yet Jackson projects a greater NFL potential upside than Mayfield, as can be seen in the distributions in Table 3.

Another interesting case is Graham Harrell and Patrick Mahomes. We mentioned earlier that Harrell was one of the QBs in Mahomes’ similarity set, but since we only look backward the reverse wasn’t true. Harrell and Mahomes were in the 2009 and 2017 draft classes respectively. Harrell and



Mahomes were both “air raid” QBs at Texas Tech. Their college career performance curves are shown in Figure 6. They are remarkably similar, and their final ratings are nearly identical. Even more interesting, is how similar they are in every passing Elo rating; yards per game (YPG), yards per completion (YPC), yards per attempt (YPA), NCAA pass efficiency (Eff), touchdowns (TD), completion percentage (PCT), and interceptions (INT).

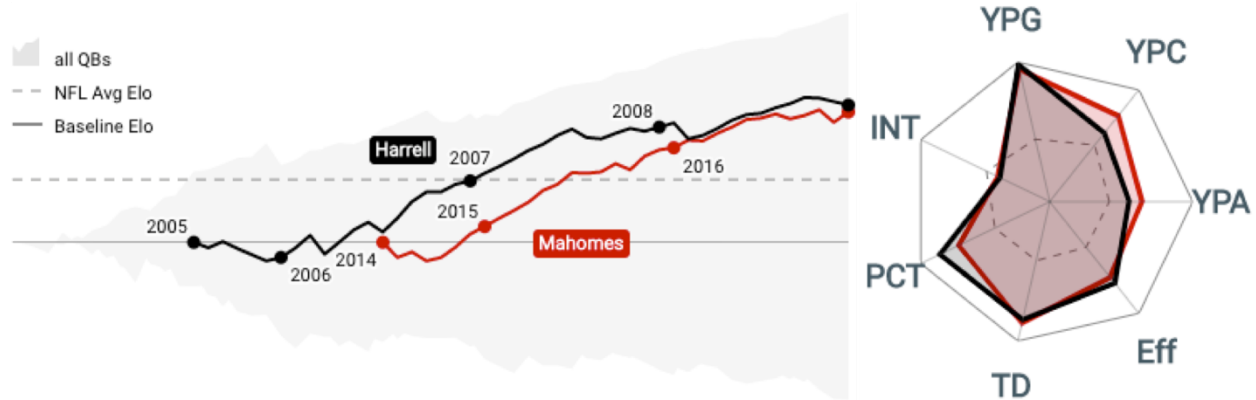


Figure 6 – Comparison of the career performance curves of Graham Harrell (black) and Patrick Mahomes (red) along with a spider plot comparing their final ratings in each of our evaluated passing statistics (right)

Harrell and Mahomes may look similar from their Elo ratings, but their projections are significantly different, as can be seen in Table 7. The first two rows show their actual projections. Harrell did not project well in 2009. The NFL has changed significantly since then. The third row shows how Harrell’s projection would change if he had been in the 2017 draft. This shows how our projections evolve with the times. NFL offenses have become more pass prolific and there are now more QBs that had similar college career curves to Harrell. Harrell’s 2017 projection is much better. He likely would have been drafted, even though he wasn’t in 2009. His potential still doesn’t match that of Mahomes, but it evolves with the league.

Table 7 – Comparing the projections of Graham Harrell, Patrick Mahomes, and a 2017 Graham Harrell

Quarterback	Similar QBs			Projection				Distribution (AV/year)
	NFL	All	NFL/All	Practice	Backup	Starter	Franchise	
Graham Harrell	3	5	60.0%	77.1%	22.7%	0.2%	0.0%	
Patrick Mahomes	28	50	56.0%	28.1%	20.1%	16.2%	30.9%	
Graham Harrell (2017)	24	32	75.0%	40.7%	15.7%	14.0%	25.3%	

## 6. Conclusion

Evaluating the NFL potential of college football players is difficult; even the most seasoned professional makes mistakes. There are only a handful of franchise players in each draft class. Making the right draft decisions can pay dividends for years to come. We find that a player’s journey to reach his final college Elo rating is a significant factor in projecting NFL potential. Our player Elo ratings ensure physical attributes and athleticism translate to on-field performance. We



use the actual NFL performance value (AV) of players with similar college career ratings curves to generate a predicted distribution of future potential. Combined with the scouting tools already available, our projections can help make draft selections less of a gamble.

## 7. Acknowledgements

I would like to acknowledge DeAngelo Haslam, Matt O'Donnell, Eric Reidelbach, and Taylor Corbett for their contributions to this effort.

## References

- [1] Lewis, M. and Leach, M. (2019) Michael Lewis – Going Deep with Mike Leach. 2019 Sloan Sports Analytics Conference.
- [2] BCS Group (2012) How to Select the Four Best Teams to Compete for the College Football National Championship. Accessed: November 9, 2019. [https://collegefootballplayoff.com/documents/2017/10/20/CFP\\_Selection\\_Committee\\_Protocol.pdf](https://collegefootballplayoff.com/documents/2017/10/20/CFP_Selection_Committee_Protocol.pdf)
- [3] Elo, A. E. (1978) The Rating of Chess Players, Past and Present, New York: Arco Publishing
- [4] College Football Stats and History. 2019, Sports Reference. Accessed: November 2019. <https://www.sports-reference.com/cfb/>
- [5] Mohebbi, M., Vanderkam, D., Kodysh, J., Schonberger, R., Choi, H., Kumar, S. (2011) Google Correlate Whitepaper. Accessed: November 9, 2019. <https://www.google.com/trends/correlate/whitepaper.pdf>
- [6] Approximate Value: Methodology. 2019, Sports Reference. Accessed: November 10, 2019. <https://www.sports-reference.com/blog/approximate-value-methodology/>
- [7] Lehman, W. B. (2019) DraftGeM: A New Tool to Compare College and Pro Football Players. Medium. Accessed: November 16, 2019. <https://medium.com/sportsraid/draftgem-a-new-tool-to-compare-college-and-pro-football-players-2b1d1432d1fc>