NBA Scoring Success

Elliot Buera, Eddy Yoon, Leo Liang, Robert Wetten

2025-04-28

Introduction

Different types of shots in Basketball are seen due to various game play scenarios that occur during basketball games. Following our initial research and observation of NBA basketball matches, we became curious to know whether overall shot success rates differ based on the type of shot. In this report, we found the most important factors that determine shot success in the NBA from our chosen shot types. We also found the top ten NBA players with the highest scoring rates for each shot type chosen.

We were interested in conducting this research because from our initial intuition, we speculated that shot types do not have the same success rate. Knowing the success rate of shot types and the success rate of individual players can potentially help NBA teams determine which shot types to attempt, determine which offensive and defensive schemes to attempt to maximize or minimize intended shot types and determine which players to play and shoot.

Our analysis focused on two shot types: Jump Shots and Layups. Our results showed that both shot types have similar factors that significantly influence shot success rates. These factors are shot distance, urgency and time and game pressure. For the top ten players for both shot types, the majority are those who play in the Center position and are some of the tallest players in the NBA.

Data

We conducted our research using the hoopR package in R. This package contains play-by-play data for all many men's basketball games, both college, and NBA. For our purposes, we were interested in analyzing the play by play data in the 2024-2025 NBA season where each row is a play that occurs during a game. This package has an abundant amount of data and is continuously updated daily as games take place. Examples of the data columns include: the type of play, text describing the play, time and period of the game, game scores for home and away teams and positional data of the play.

Since our research focuses on shots, we removed all rows that were plays not involving shots. Continuing our research on shots, we made a bar chart that highlights the different types of shots in Figure 1.

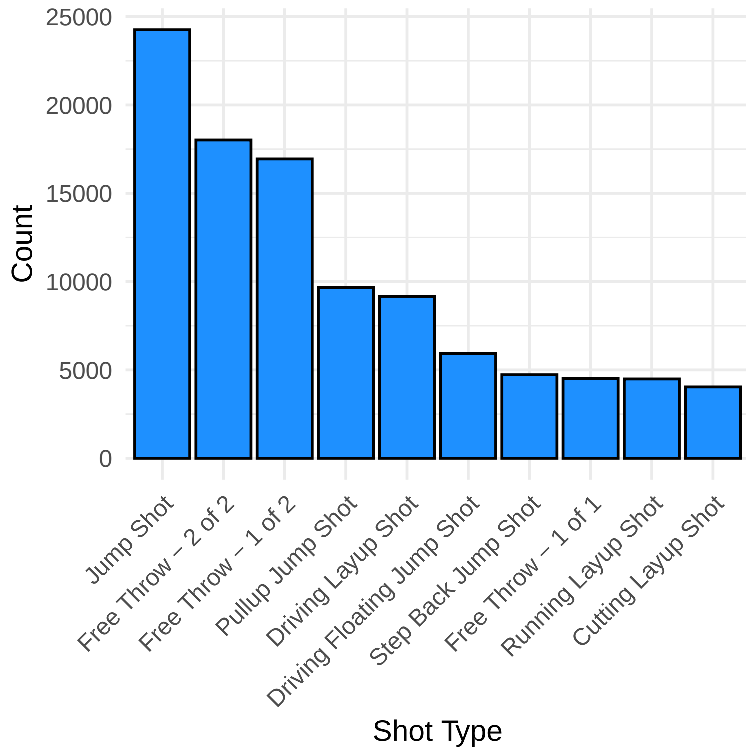

Figure 1: Top 10 most frequent shot types for made shots in the NBA. The bar chart shows that many shots take the lable of being a jump shot or a layup, which motivated us to focus on these shot types.

The Top 10 Ways NBA Players Score bar chart above shows the counts of all types of successfully made shots. We discovered that there are two main types of shots that could be merged into broader categories: Jump Shots and Layup Shots. Once creating two new data frames filtered to just be shot descriptions containing "Layup" and shot descriptions containing "Jump Shot," we decided to explore whether these variables made sense.

From our understanding of basketball, we noticed that Jump shots can be of any distance but Layup shots must be a close range shot. We decided to further investigate whether our newly merged Layup shot category contains shots that are only of close range. Our findings (shown in the left panel of Figure 2), show that there are shots that are outside of the usual Layup shot range included when. This may indicate that some plays are mislabeled as layup shots or are not the traditional layup shots that we are interested in. We decided to remove rows that are layup shots and have distances that are too far from the basket and plot our Heat Map of Layup Shots again. Figure 2 displays both plots and proves our suspicion correct. Many shots were classified as layups despite being very far from the basket.

The new Heat Map includes arbitrary Layup Shot distances that we selected. The Endline axis now has a shorter range than the previous heat map.

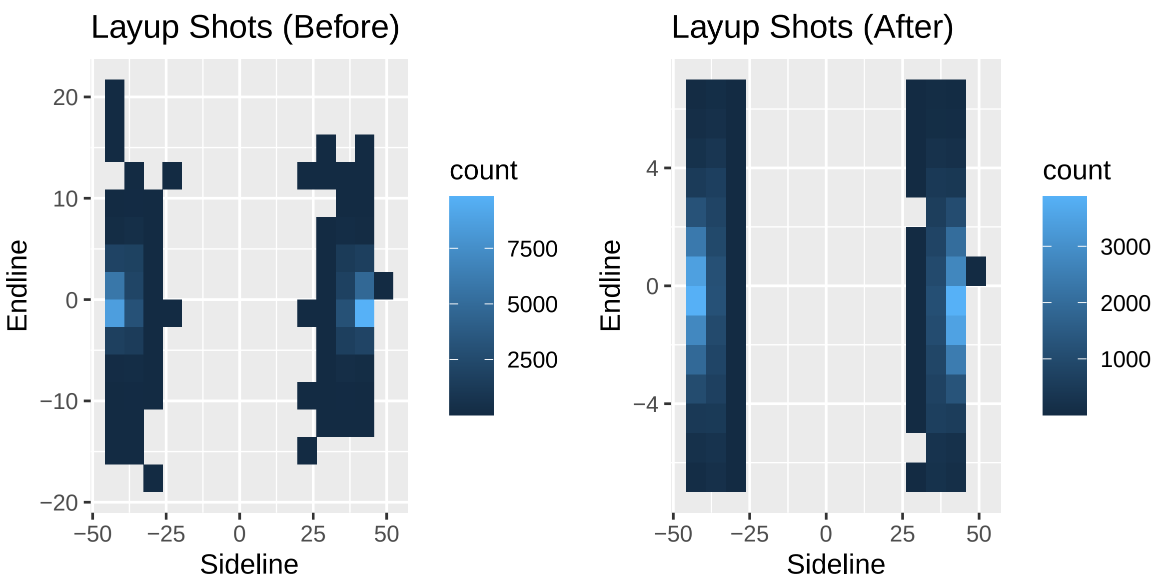

Figure 2: Layup shots heatmap before (left) and after (right) adjusting to ensure that layups are taking place 'close' to the basket. We can see that many shots are filtered out, which could be key for finding curitl results.

Finally, after observing our preliminary EDA, our goal with the data was to predict whether or not a score would be made when players were taking jumpers and layups. For this reason, we decided to conclude by visualizing the distribution of made shots across both layups and jump shots. Figure 3 shows the results.

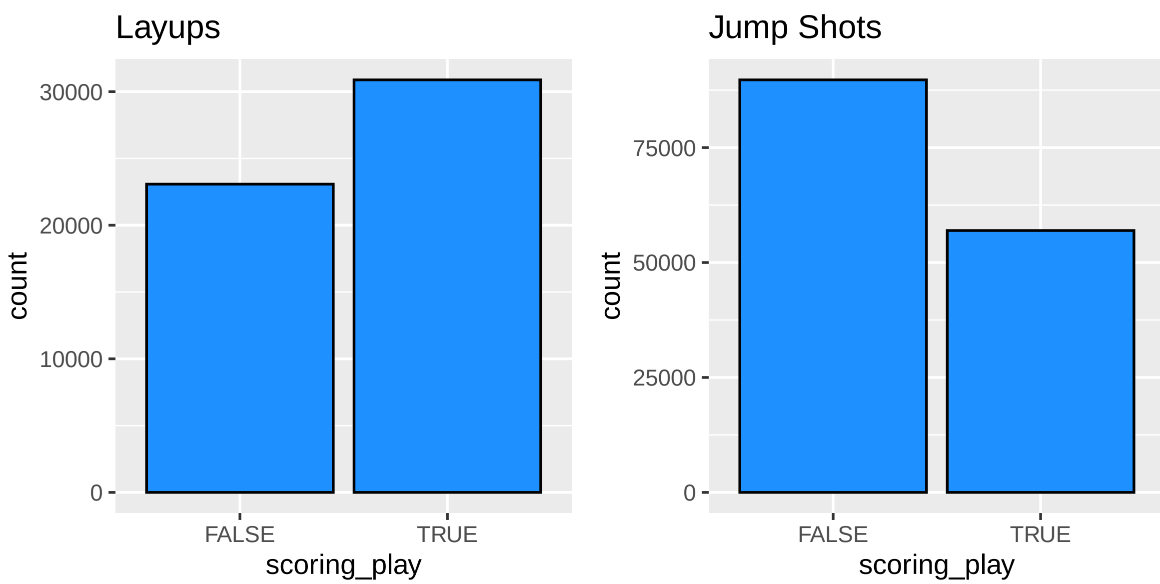

Figure 3: Distribution of made shots for both layups (left) and jumpshots (right). We see that the success rate for layups is much greater than the success rate of jump shots. Additionally, there is some imbalance in both distributions

Methods

Layup Model

After already loading the 2025 NBA play-by-play data and 2025 player box-score data, we then filtered the play-by-play records to retain only layup attempts. Each layup attempt was joined to its corresponding box-score line by matching on game identifier and player identifier. From this combined dataset, we derived the following variables for each shot: a binary success indicator, Euclidean shooting distance from the basket, seconds remaining in the quarter, a flag for whether the shooter was in the starting lineup, the shooter's season field-goal percentage, the shooter's points scored in that game, and the quarter in which the attempt occurred. We standardized all continuous predictors to have mean zero and unit variance so that their estimated effects would be directly comparable and the model would converge more reliably.

To model layup success, we fit a logistic mixed-effects regression using the lme4 package's glmer function with a logit link. In this framework, the log-odds of making a layup are modeled as a linear combination of the standardized shooting distance, standardized seconds remaining, starter status, standardized field-goal percentage, standardized points, and quarter, plus a player-level random intercept to capture each shooter's baseline ability. We assume that, conditional on these random intercepts (which follow a bell-shaped distribution centered at zero), individual shot outcomes are independent and follow a Bernoulli distribution. We used the Bobyqa optimizer to ensure stable convergence given the hierarchical structure and added covariates.

We chose logistic regression because our outcome is binary (make vs. miss) and the logit link naturally maps probabilities to the real line. The mixed-effects extension is appropriate because shooter skill varies systematically across players; by including a random intercept, we pool information across athletes, stabilizing estimates for those with fewer attempts while avoiding the overfitting that would result from fitting a separate fixed effect for every player. Incorporating box-score variables like field-goal percentage and points scored adds useful context about each player's shooting ability and game performance.

To evaluate model performance, we compare the mixed-effects model's AIC to those from a simpler logistic regression without random intercepts. Furthermore, to explain the certainty of our model's estimates, we will model a 95% confidence interval for each predictor to (hopefully) ensure that many do not contain 0 within the interval.

Jumpshot Model

We will use a logistic mixed-effects regression model to predict whether an NBA jump shot results in a made basket (scoring_play = 1). Fixed effects include shot difficulty and game context variables such as shooting distance, time remaining in the quarter, field goal percentage, three-point percentage, total points, starter status, and period number. A random intercept for each player (athlete_id_1) accounts for player-specific differences in baseline shooting ability. The response variable follows a Bernoulli distribution, and random effects are assumed to follow a normal distribution. Although individual shots are not truly independent, conditional independence given the model structure is a reasonable approximation. In our circumstances, and similarly for the layup model, a mixed-effects model is appropriate because the data is hierarchically structured, with many shots taken by the same players. Accounting for this structure avoids underestimating standard errors. Fixed effects like shooting distance and field goal percentage are logical determinants of shot success.

In order to evaluate the model's performance, we will again compare the mixed-effects model's performance to a standard logistic regression model without random effects and check for 0s in the 95% confidence interval of our predictors.

Results

Layup Results

After fitting both the mixed effects logistic regression model and the plain logistic regression model, we compared the AIC values to analyze our model's predictive powers. Table 1 shows the results from our test.

Table 1: AIC comparison between simple logistc regression model and mixed effect model (layup)

	df	AIC
layup_model_mixed	12	65104.07
layup_model_plain	11	65374.00

Model Comparison:

We compared a mixed-effects logistic regression model to a normal logistic regression model with identical fixed effects to see if adding random effects had a significant impact on layup success. The mixed-effects model had an AIC of 65,104 which is lower than the plain logistic regression model's AIC of 65374, indicating that accounting for shooter-level heterogeneity improves model fit despite the extra parameter.

Fixed Effects:

Table 2: Fixed effect estimates from our mixed logistic regression model (layups)

Term	Estimate	Std. Error	Z value	P value	CI lower	CI upper
(Intercept)	0.274	0.031	8.918	0.000	0.214	0.334
scale(shooting_distance)	0.419	0.010	42.117	0.000	0.400	0.439
scale(start_quarter_seconds_remaining)	-0.035	0.010	-3.628	0.000	-0.054	-0.016
starterTRUE	0.067	0.029	2.353	0.019	0.011	0.123
scale(field_goal_pct)	0.867	0.014	60.813	0.000	0.839	0.895
scale(points)	-0.087	0.015	-5.712	0.000	-0.117	-0.057
factor(period_number)2	0.014	0.027	0.534	0.594	-0.038	0.067
factor(period_number)3	0.004	0.027	0.143	0.886	-0.049	0.056
factor(period_number)4	0.019	0.028	0.701	0.484	-0.035	0.073
factor(period_number)5	-0.212	0.134	-1.585	0.113	-0.475	0.050
factor(period_number)6	0.248	0.564	0.439	0.661	-0.858	1.354

Table 2 displays our fixed effects estimates with their uncertainty estimates. Looking at our fixed-effect estimates, shooting distance shows a strong positive association with layup success that is statistically significant (95% CI [0.400, 0.439]; $p < 0.001$). Specifically, a one-standard-deviation increase in distance corresponds to around a 42% increase in the odds of making the shot, suggesting that longer layups tend to be attempted by higher-skill finishers. The amount of time remaining in a quarter is also statistically significant (95% CI [-0.054, -0.016]; $p < 0.001$). Taking the shot earlier in the quarter (one-SD more seconds remaining) reduces the odds by about 3.5%, implying that late-clock situations yield easier layup opportunities. Starter status also matters: starters have roughly 7% higher odds of converting a layup than non-starters (95% CI [0.011, 0.123]; $p = 0.019$). Season field-goal percentage is our strongest predictor where each one-SD increase more than doubles layup odds (around 86.7%; 95% CI [0.839, 0.895]; $p < 0.001$), while scoring more points in the same game is associated with about an 8% drop in odds (95% CI [-0.117, -0.057]; $p < 0.001$), perhaps reflecting fatigue or tighter defense. None of the later quarters (2–4) differ meaningfully from the first, and although period 5 and 6 (overtime) shows larger negative estimates than quarters 2–4, its wide 95% CIs and larger p-values mean we cannot rule out chance. These findings, with 95% confidence intervals around each effect, indicate that individual skill (as captured by season shooting percentage and starter role) and situational factors (distance and late-quarter timing) drive layup success far more than the nominal quarter, guiding coaches toward emphasizing player shooting talent and end-of-clock plays rather than worrying about which period the shot occurs in. We account for uncertainty in these inferences through 95% confidence intervals for each coefficient and by validating the model's predictive performance out-of-sample, giving us confidence that these effects are robust and substantively meaningful in guiding shot-selection strategies.

The distribution of player random intercepts from the mixed model is displayed in Figure 4.

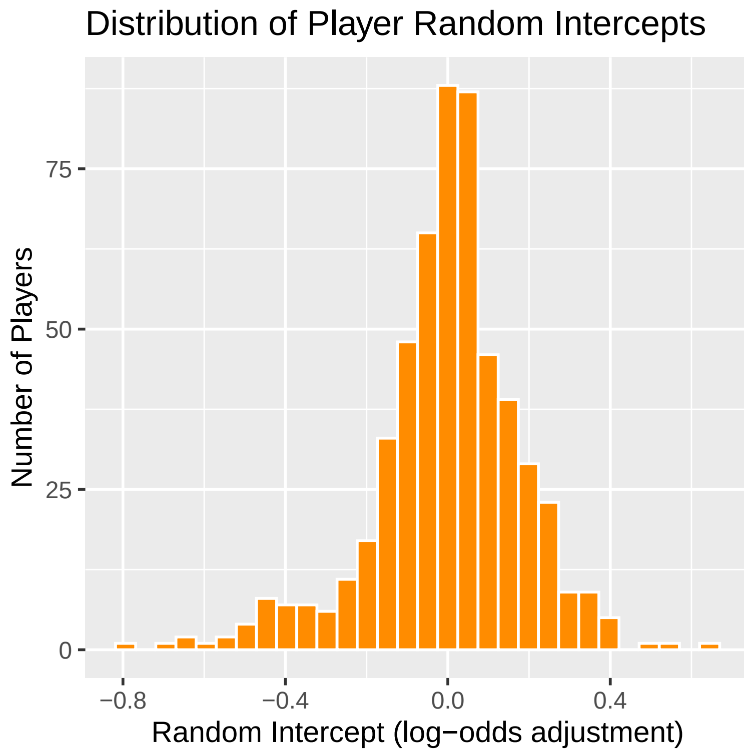

Figure 4: The histogram of player random intercepts shows that most athletes' adjustments cluster tightly around zero—indicating average baseline layup ability—while a tail of players achieve intercepts as high as above +0.5 log-odds (well above average) or as low as –0.8 (well below average).

The spread in the histogram confirms that, even after accounting for distance, clock, and shooting talent, there remain substantial between-player differences in finishing skill. By explicitly modeling these random intercepts, we ensure that our estimates of situational factors (like distance and quarter timing) are not biased by individual shooters' overall ability levels.

Jumpshot Results

For the jump shot results, we followed a very similar process. Table 3 displays the AIC values between our mixed effects model and our plain comparison model.

Table 3: AIC comparison between simple logistc regression model and mixed effect model(jumpshot)

	df	AIC
model_mixed	13	172913.4
model_plain	12	172940.7

Table 3 shows that the mixed effects model has a lower AIC again. This shows strong evidence that accounting for player-specific variability with a random intercept improves model fit.

Table 4 displays the fixed effects from the mixed model.

Table 4: Fixed effect estimates from our mixed logistic regression model (jump shots)

Term	Estimate	Std. Error	Z value	P value	CI lower	CI upper
(Intercept)	-0.470	0.017	-28.161	0.000	-0.502	-0.437
scale(shooting_distance)	0.112	0.006	19.058	0.000	0.101	0.124
scale(start_quarter_seconds_remaining)	0.050	0.006	8.388	0.000	0.038	0.062
starterTRUE	-0.016	0.016	-1.039	0.299	-0.048	0.015
scale(field_goal_pct)	0.517	0.009	57.142	0.000	0.499	0.535
scale(three_point_pct)	0.358	0.007	48.374	0.000	0.344	0.373
scale(points)	0.049	0.008	5.866	0.000	0.033	0.066
factor(period_number)2	-0.058	0.016	-3.550	0.000	-0.089	-0.026
factor(period_number)3	-0.017	0.016	-1.034	0.301	-0.048	0.015
factor(period_number)4	-0.105	0.017	-6.296	0.000	-0.138	-0.072
factor(period_number)5	-0.178	0.084	-2.122	0.034	-0.343	-0.014
factor(period_number)6	0.163	0.322	0.507	0.612	-0.467	0.794

Several fixed effects were associated with jump shot success. Shooting distance had a small positive effect, suggesting that longer jump shots, likely those that are more open, have slightly higher success rates when controlling for other factors. Shots taken earlier in the quarter were more likely to be made, likely because late-clock attempts are lower quality. Player performance metrics had strong effects, with higher field goal percentage, three point percentage, and total points increasing the likelihood of a made shot, consistent with hot hand behavior. Starter status was not significantly related to shot success after accounting for player efficiency. Shots taken in later periods, especially the fourth quarter, were less successful, likely due to fatigue or defensive pressure. Overall, both player efficiency and game context emerged as key drivers of jump shot success.

Each fixed effect coefficient is accompanied by a 95% confidence interval. All major effects (shooting distance, time remaining, field goal percentage, three-point percentage, points) have tight confidence intervals that do not cross zero, reinforcing that these predictors are reliably associated with jump shot success. In contrast, non-significant predictors like starter status and period 3 have confidence intervals that contain zero, so we have some uncertainty about their effects.

The distributions of the player's random effect intercepts is shown in Figure 5.

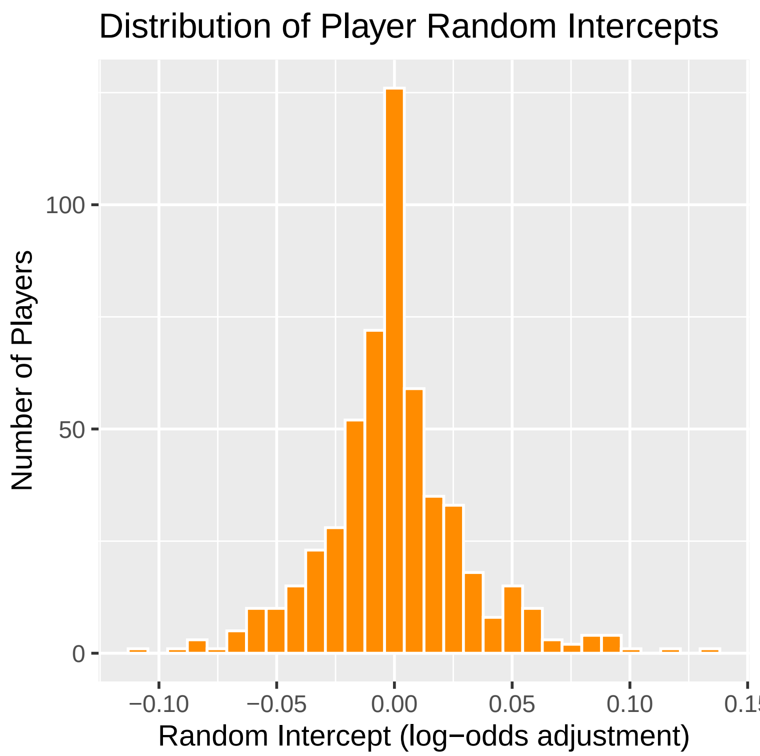

Figure 5: The distribution of player random intercepts is centered around zero with a moderate spread, suggesting that while individual players do differ in jump shot success after accounting for context, these differences are generally small. A few players exhibit noticeably higher or lower baseline success rates, indicating a minority of especially skilled or less effective jump shooters.

Similarly to the layups model, the distribution of random effects is tightly centered around 0 with minimal spread (approximately normally distributed).

Discussion

Overall, our main findings from this report indicate that in both the layups case and the jump shots case, adding a mixed effect for player-specific intercepts improved the model fit when compared to simple logistic regression. For layup success, despite questioning whether shot distance would take precedence, we still found that the distance the player was away from the basket played a positive role in determining where shots would be made. More successful players tend to make layups from longer distances. The same is true for jump shots. Additionally, we found the for both layups and jump shots, the quarter/period number does not play a crucial role in determining shot making, but that there does seem to be some negative trend as players get later in the game, they seem to get more fatigued.

Some limitations from our model is that it does not include a mixed effect for defenders, which would be very crucial in determining a player's success on offense, as not all defenders are created equal. We believe that scoring on James Harden is likely much easier than scoring on say, Victor Wembanyama. Additionally, due to the nature of our model, it fails to differentiate shot types such as "Pullup Jump Shot" and "Fade Away Jump Shot" since these shots are pulled together (the same goes for layups of all types). The limitation is that we don't account for making harder shots, which is problematic because it treats all shot types the same.

In the future, we wish to address some of these limitations. Specifically, the hoopR package contains vast data on all sorts of metrics. We believe that we could apply some defensive metrics to our modeling to account for mixed effects from different defensive teams our defensive players. Specifically, we imagine that there must be some negative effect to making layups when a player is being guarded by an incredible blocker. Furthermore, our model only considers data from the 2024-2025 season. We believe that conducting this research on games from in the past, could lead to results that show how different metrics have evolved over time. For example, for jump shots, we believe there is possibly an effect that as players began to take more jump shots, they've gotten more skilled in doing so, and this could lead to an effect showing that shooting distance was more of a consideration in the past. In a follow up, we suggest that statisticians attempt these approaches.

Overall, we have gained some important insights and have contributed to learning what contributes to *NBA Scoring Success*.