What if we’ve been arguing about the NBA GOAT using the wrong data all along?

Traditional metrics have convinced us that greatness can be reduced to points, efficiency, or rings. But basketball is ultimately about winning. **Winning** means impacting the outcome of every game, not just stacking stats. **Winning** requires influencing your teammates and outmaneuvering your opponents, game after game. **Winning** is what separates legends from the rest — and that’s exactly what this new metric measures.

Inspired by the **ELO system from chess**, one of the most respected and widely used rating systems in the world, this metric applies principles that have proven reliable for decades in a completely different competitive context to measure a player’s impact with precision and consistency.

One of its most powerful advantages is that it allows us to **compare players from completely different eras**, even if they never faced each other. Most traditional metrics fail here because they ignore the quality of the opposition, treating every point or win as equal. Michael Jordan is often accused of playing against “plumbers,” while people mock LeBron James for playing in a soft era of “TikTokers.” This metric, however, **accounts not only for the level of the opponent but also the strength of the player’s own team**, providing a fair way to evaluate dominance across generations and finally ask the hard question: **who truly rises above the rest, no matter the era?**

Another key feature is that it **rewards overperformance against strong opponents and penalizes underperformance against weak ones**. Unlike traditional box-score metrics, it recognizes that a 30-point blowout win against a weak team isn’t the same as leading your squad to victory against the league’s best, while losing to a weaker opponent can’t be ignored. By capturing both successes and failures relative to the quality of competition, the metric emphasizes accountability, consistency, and real impact. Those who consistently overperform under pressure rise to the top, while players who occasionally shine but underperform in critical moments are exposed — mirroring how chess ratings balance wins and losses to measure true skill.

By adapting ELO to basketball, we can finally **quantify greatness in a way that respects context, competition, and true influence on the court**. It’s not just about points or rings — it’s about winning consistently against the best and avoiding costly losses against the worst. This metric provides a fresh lens for the GOAT debate, one grounded in logic, data, and competitive reality.

Before diving into basketball, it’s worth recalling how the **ELO system works in chess**. The ELO system in chess is a simple yet powerful way to quantify a player’s skill relative to their opponents. Each player starts with a base rating, and after every game, their score is adjusted based on the result and the rating of the opponent. **Beating a higher-rated opponent increases your rating more than beating a lower-rated one**, while losing to a weaker player results in a larger decrease. Over time, this dynamic system produces a numerical rating that reflects not just wins and losses, but the **context and difficulty of each match**, allowing players from different eras or tournaments to be compared on a common scale.

After every game, the rating is adjusted based on the result and the relative rating of the opponent. The formula is:

Where is the current rating, the updated rating, the sensitivity factor, the actual result (1 for win, 0.5 for draw, 0 for loss), and the expected result calculated using the opponent’s rating:

This system has proven extremely effective at **measuring skill dynamically**, taking into account both victories and defeats relative to the strength of the opposition.

With this foundation in mind, the first challenge in adapting this system to basketball is that it is a **team sport**, and we aim to measure the **performance of an individual player** rather than the team as a whole. In chess, each game has a single winner and loser, but in basketball, multiple players contribute simultaneously and diversely to a single outcome.

To address this, we first need **to define the rating of a team based on the ratings of its individual players**. A natural approach is to consider the team rating as the **weighted average of the ratings of all players on the court**, where weights reflect minutes played. This makes intuitive and mathematical sense because not all players contribute equally to the outcome of a game: a starter playing 36 minutes has far more influence than a bench player with 5 minutes on the court. Weighting by minutes ensures that the team rating **accurately reflects the actual influence of each player**, rather than treating all participants as equal, which would distort the expected outcome and undermine the reliability of the ELO updates for individual players. Formally,

Where:

* is the rating of the team
* is the current rating of player , before playing the game
* is the number of minutes player played in the game
* is the total number of players who participated in the game

One important aspect of this metric is that it **incorporates the actual minutes each player spends on the court during the game**, which cannot be known before the match begins. This means that in-game events, such as injuries or unexpected substitutions, naturally affect the player’s contribution and rating adjustments. In other words, the metric captures a player’s **true impact on the outcome** rather than an estimate based on pre-game expectations. This makes the rating more accurate and reflective of real performance, but it also means that it cannot fully predict performance ahead of time—it measures impact **ex post** rather than forecasting it.

Before moving forward, let’s illustrate the calculation above using a historic game: **the Game 6 of the 1985 NBA Finals matchup between the Los Angeles Lakers and the Boston Celtics**, featuring legends like Magic Johnson, Kareem Abdul-Jabbar and Larry Bird. In the following tables, the minutes played by each player and their ranking before the game are shown for the two teams.

|  |  |  |
| --- | --- | --- |
| Player | Minutes played | ELO ranking |
| Larry Bird | 47 | 1550.33 |
| Danny Ainge | 43 | 1131.05 |
| Dennis Johnson | 43 | 1396.36 |
| Kevin McHale | 42 | 1222.15 |
| Robert Parish | 39 | 1235.19 |
| Scott Wedman | 15 | 825.17 |
| Greg Kite | 11 | 984.96 |

|  |  |  |
| --- | --- | --- |
| Player | Minutes played | ELO ranking |
| James Worthy | 45 | 1107.41 |
| Magic Johnson | 36 | 1459.90 |
| Byron Scott | 35 | 1192.98 |
| Kareem Abdul-Jabbar | 35 | 1689.86 |
| Michael Cooper | 30 | 1312.92 |
| Kurt Rambis | 26 | 1211.96 |
| Mitch Kupchak | 20 | 1164.60 |
| Bob McAdoo | 10 | 1019.67 |
| Mike McGee | 3 | 1132.36 |

Applying the formula above, the team ELO ranking for the los Angeles Lakers was 1296.14, and for the Boston Celtics, it was 1267.75.

Once we have defined the ratings of the two teams, we can calculate the **expected outcome** for a matchup using the classic ELO formula, applied to team ratings:

Where and are the ratings of Team A and Team B, respectively.

Let’s give some intuition on the formula. The expected score essentially represents the **probability that a team will win** based on its rating relative to the opponent. If two teams have identical ratings, , meaning each team has an equal chance of winning. As the rating difference grows, shifts closer to 1 for the stronger team and closer to 0 for the weaker team, reflecting an increasingly lopsided expected outcome.

In other words, captures **how surprising a result would be**: a win against a much stronger team counts more, while a loss to a much weaker team counts more negatively. This is exactly why ELO naturally rewards overperformance and penalizes underperformance, making it a dynamic and context-sensitive measure of success. **For every 400-point difference in rating, the higher-rated team’s expected win probability is 10 times more than the lower-rated team’s expected win probability**, illustrating how strongly the formula weighs large differences in skill.

Continuing with the previous example of the Finals game between the Los Angeles Lakers and the Boston Celtics, in which we had calculated Boston’s rating as 1267.75 and Los Angeles’ rating as 1296.14, the formula above tells us that the probability of Boston winning was 45.9%, which gives 54.1% chances of Los Angeles winning. As can be seen, higher rating means greater win expectancy.

Once the expected outcome is calculated, we update each player’s individual rating using a modified ELO formula:

Where:

* is the current rating of player , before the game
* is the updated rating after the game
* is the scaling factor controlling sensitivity: **16 for regular season games** and **32 for playoff games**, reflecting the increased importance of postseason contests
* is the number of minutes player played in the game
* is the actual result of the team (1 for win, 0 for loss)
* is the expected outcome of the team based on team ratings

This formula captures a player’s true impact on the game in a dynamic, context-sensitive way. The term is always positive when the team wins (because the team’s expected result is always less than 1), which means the player’s rating increases. Conversely, if the team loses, the term is negative (because the expected result is always greater than 0), and the player’s rating decreases.

**It’s fundamental to note that if you don’t win, your rating does not improve**, no matter how spectacular your performance is. For example, Michael Jordan’s **63-point game against the Boston Celtics in the 1986 playoffs** or LeBron James’ **51-point game against the Golden State Warriors in the 2018 Finals** both resulted in **ELO decreases**. And that’s exactly how it should be: in basketball, **winning is all that matters**. Consequently, your rating improves if, and only if, your team wins.

That being said, observing the formula, the change in a player’s rating depends on **three main components**: the difference between the actual and expected outcome, the player’s minutes played, and the scaling factor .

1.- The **difference between actual and expected outcome measures overperformance or underperformance relative to expectations**. If a team wins against a stronger opponent, is greater than if they beat a weaker one, leading to a greater rating increase. Conversely, if the team loses to a weaker opponent, the term is negative, causing a decrease. This is the core component that reflects whether a player’s team—and by extension, the player themselves—exceeded or fell short of what was expected.

| **Result** | **vs. Strong Opponent** | **vs. Weak Opponent** |
| --- | --- | --- |
| **Win** | Rating increases ‘a lot’ | Rating increases ‘a little’ |
| **Loss** | Rating decreases ‘a little’ | Rating decreases ‘a lot’ |

2.- **Minutes played ()**. Not all players have the same influence in a game. By scaling the rating change according to minutes played—normalized to a 36-minute standard—the formula ensures that players who spend more time on the court have a proportionally greater impact on their rating. A starter playing the bulk of the game will see larger rating adjustments than a bench player with limited minutes, even if both experienced the same team result. Consequently, a loss penalizes starters more than bench players, and a win rewards starters more heavily as well, reflecting their greater contribution to the game outcome.

**3. Scaling factor ().** The factor adjusts the magnitude of rating changes based on **game importance**. In this system, regular season games use , while playoff games use , reflecting the higher stakes and pressure of postseason contests. A bigger means that playoff performances move ratings more dramatically, ensuring that clutch performances or costly failures in high-stakes games are weighed appropriately. The use of these values of are a common practice in chess.

Together, these three components determine **how much a player’s rating rises or falls after each game**, combining performance, opportunity, and context into a single, interpretable metric.