Jacob Jannicelli

# Pitcher Rating

**Abstract**

This project develops *Pitcher Rating* (pRating), a comprehensive sabermetric designed to quantify total pitcher performance in a single, interpretable value. Modern baseball statistics often fragment pitching into isolated traits, making holistic comparison difficult. pRating integrates five major performance dimensions — command, dominance, efficiency, contact management, and run prevention — using Location+, pitches per 9 innings, strikeouts per 9, expected total bases per 9, and expected ERA. These components are weighted and normalized to reflect both short-term and long-term pitching ability, then rescaled into pRating100, a standardized 100-point version that enhances interpretability and cross-season comparison. Empirical tests show meaningful correlations between a pitcher's pRating100 and their following season ERA, as well as moderate year-to-year stability, indicating that the metric captures persistent elements of true pitching skill. While certain limitations remain, including imperfect data joins, outlier performances, and subjective weighting, pRating demonstrates clear analytical value for scouts, coaches, front offices, and fans by condensing multiple aspects of performance into one cohesive, predictive metric.

## History

Baseball statistics have become increasingly fragmented over time, each often measuring only one isolated trait. Pitcher Rating (or pRating) brings together several of these smaller, fragmented metrics into one overall performance metric. pRating can be applied to both a single-game outing, as well as season- or career-long performances, making it possible to compare

players to one another in the short- or long-term. As many sabermetrics do, pRating aims to isolate the individual pitcher's ability from other external factors, such as game situation, luck, and team defensive strength.

The most similar historical statistic to pRating is Game Score, a formulaic metric devised by Bill James in the 1980s and updated later by Tom Tango. Game Score consolidates box score statistics into a calculated performance score for a pitcher's single game. The score starts at 50, adding points for positive box score stats (i.e. +1 point for a strikeout, +2 points for each inning completed after the $4^{th}$) and subtracting points for negative box score stats (i.e. -2 for each hit allowed, -1 for each walk allowed), typically ending up ranging from 0 – 100. Tom Tango's later version starts with a lower baseline, introduces a penalty for home runs allowed, and equalizes the penalty for walks and hits. Though this metric is interesting and aims to create a holistic score, the cumulative metrics used are a bit primitive (mostly "counting" stats), it doesn't account for many external factors, it penalizes relief pitchers who don't have long outings, and it is limited to single-game performance. pRating aims to address these shortcomings.

**Process and Output**

I selected 5 categories of metrics that cumulate to the overall pRating: command, dominance, efficiency, contact management, and run prevention. For command, I used a form of Location+, which is an existing sabermetric that judges a pitch on the location it was thrown in varying situations, not just if it was a ball or strike. For example, a curveball thrown below or away from the zone on a 0-2 count isn't necessarily a bad pitch, while a fastball thrown center-high in the zone on a 0-2 count isn't necessarily a good pitch. My Location+ stat has target zones for varying combinations of pitcher handedness, batter handedness, pitch type, and count rewarding pitchers for throwing pitches that are accurate to those zones and according to the

game situation. For efficiency, I used pitches per 9 innings. Getting through innings in fewer pitches is more efficient, reflective of quick and dominant outs, and allows a pitcher to throw for longer in a game. For dominance, I used strikeouts per 9 innings. Getting a strikeout is the most dominant statistic for a pitcher, relying only on themself and not their defense to get outs. For contact management, I used expected total bases per 9 innings. Expected total bases aims to remove the luck factor from a ball in play by determining if a batted ball should have been a single, double, triple, or home run based on its exit velocity, launch angle, and distance. This, added with walks, gives the total bases allowed by a pitcher in 9 innings. Finally, I used expected ERA for run prevention. Expected ERA uses similar logic as expected total bases to more accurately determine how many runs a pitcher should have given up in a 9 inning game. All of these metrics are adjusted to this 9-inning standard in order to attempt to give some comparative value between starters, relievers, and closers. These five metrics are given weights and calculated into pRating, as summarized below.

| Component | What It Captures | Weight |
|---|---|---|
| `Location+` | Command & execution | 0.11 |
| `K/9` | Ability to miss bats | 0.26 |
| `Pitches/9` | Efficiency | 0.11 |
| `xTB/9` | Quality of contact allowed | 0.16 |
| `xERA-` | Expected run prevention | 0.36 |

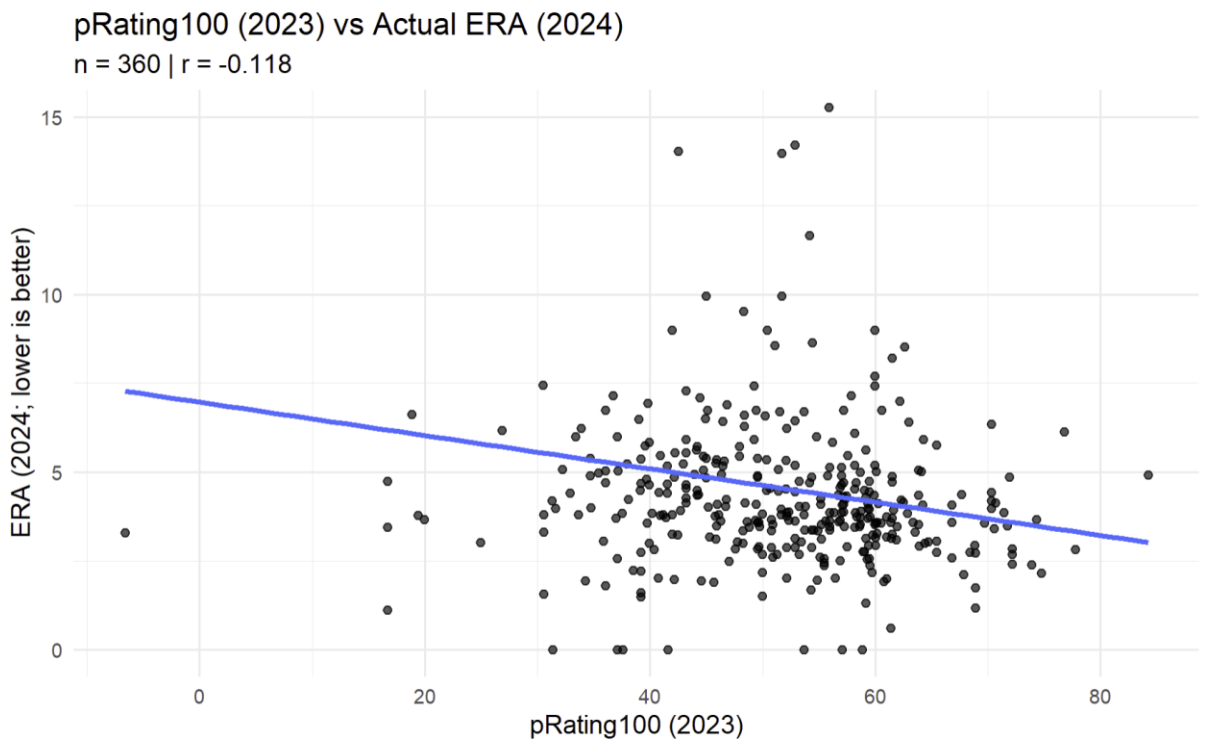
These weights were chosen subjectively according to what I thought was most relevant to true success. Location + is assuredly the least definitive metric of success, as pitchers will occasionally throw pitches away from the “optimal” zone in order to later set up a more important pitch in the count; or, to pitch away from batter’s hot zones or into their cold zones.

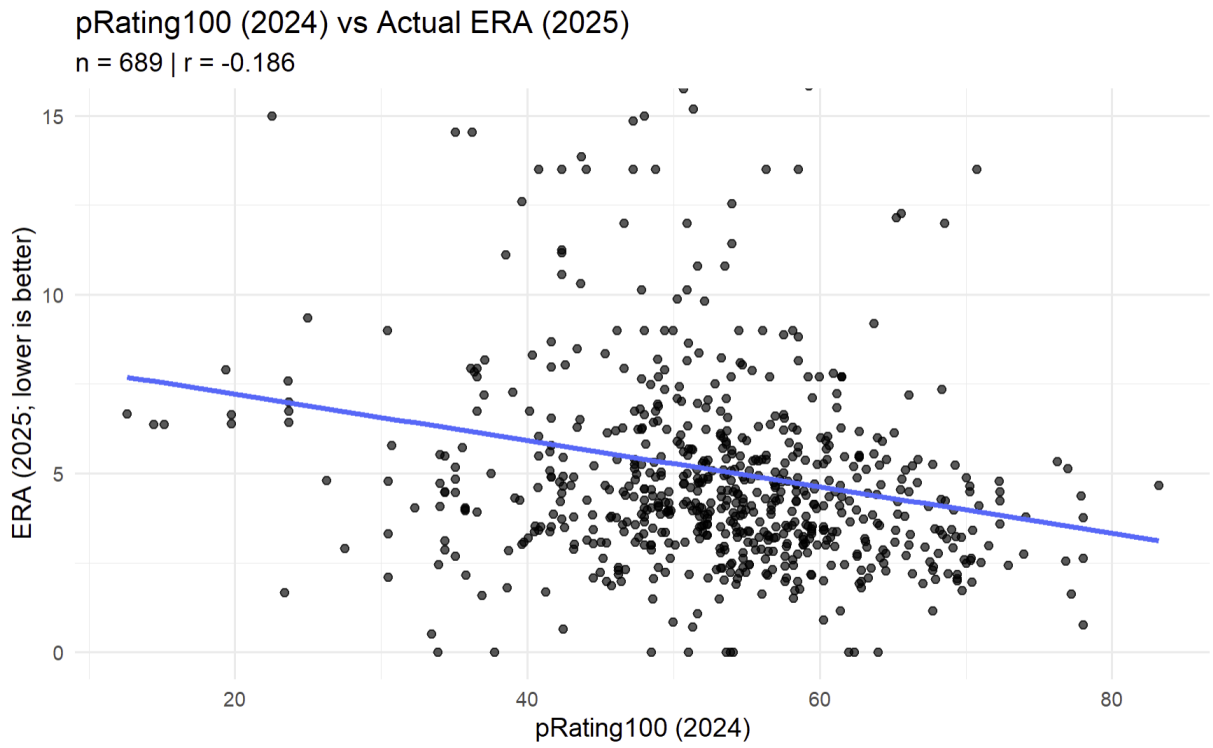
Pitches/9 also gets a lower weight, because, for example, there is no real in-game benefit to getting an out in 3 pitches versus 6, but rather more of a long-term or conceptual efficiency benefit. xTB/9, K/9, and xERA- receive more weight, as three of the best aspects of a pitcher's performance are often lack of baserunners, more strikeouts, and no runs allowed.

Additionally, pRating is transformed into pRating100, which rescales the metric into a consistent, interpretable 100-point format, where the mean score is 50 and the standard deviation is 15. This allows for better comparison and identifying above and below average players and games. Leaders for pRating100 across 2023-2025 include Shohei Ohtani (92.7), Robert Stephenson (91.6), and Edwin Diaz (87.5), while standout arms such as Mason Miller (81.4), Tarik Skubal (80.9) and Paul Skenes (78.5) make the top 15.

**Significance**

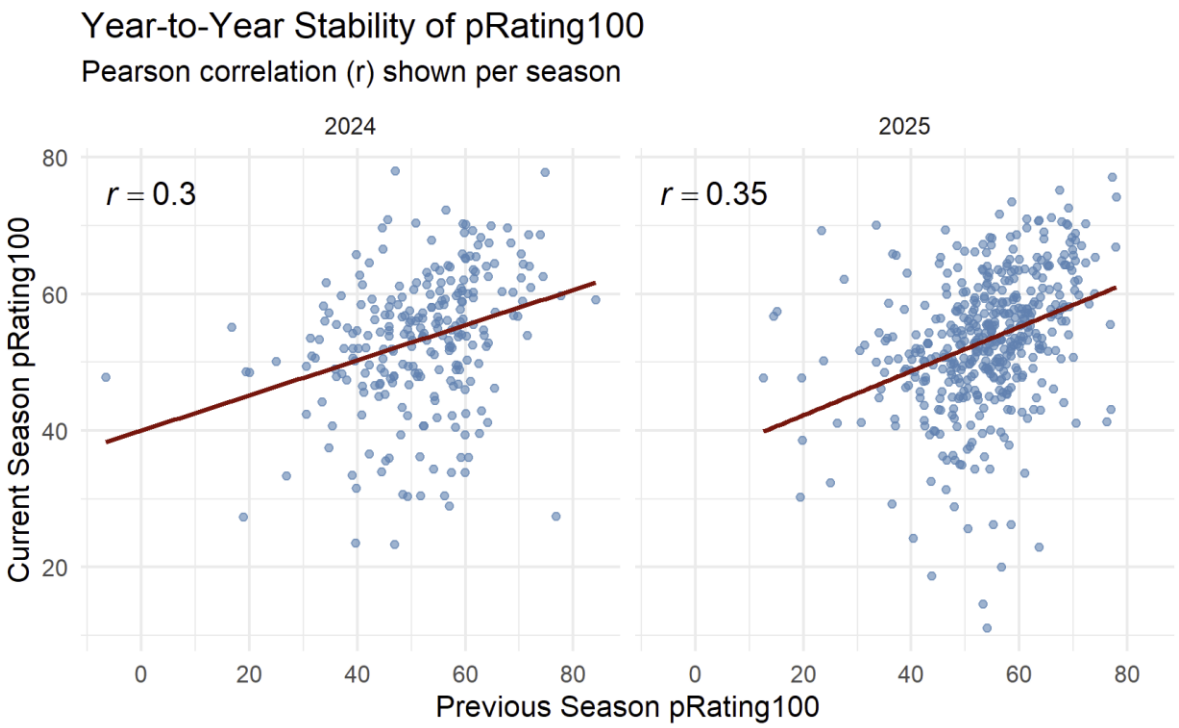
To prove that pRating100 is significant, I plotted each pitcher's pRating100 against their next year's ERA. ERA data were taken from Baseball-Reference.com and joined in with the existing yearly pitcher data for each player. It is important to note that some players' data were lost during this join (as seen in the n values), and some showed outlier ERA values even above 15, but these graphs are still as accurate as possible and tell a solid predictive story.


pRating100 (2023) vs Actual ERA (2024)
n = 360 | r = -0.118
ERA (2024; lower is better)
pRating100 (2023)
0
5
10
15
0
20
40
60
80


pRating100 (2024) vs Actual ERA (2025)
n = 689 | r = -0.186
ERA (2025; lower is better)
pRating100 (2024)
0
5
10
15
20
40
60
80

As you can see, these graphs show some significant correlations between one year's pRating100 and the next year's ERA. In other words, players with a greater pitcher rating are more likely to have a lower ERA the following year.

Additionally, I wanted to show the stability of pRating100 over time. Moderately strong correlation between season pRating100 against previous season pRating100 is shown below.




Though these plots and the vsERA plots show some promising statistical significance, I believe that the correlation values for all plots will be even stronger with more accurate data, iterative adjustments of the weights of pRating, and more sophisticated handling of external factors (i.e. adjusting for opponent difficulty or unique stadium dimensions).

## Next Steps

The next phase of development for pRating will focus on three key improvements. First, the metric's component weights will be systematically adjusted and optimized to enhance both predictive performance and year-to-year stability, replacing subjective choices with data-driven tuning. Second, outlier values — particularly extreme single-game performances and anomalous ERA outcomes — will be corrected or smoothed to ensure the dataset more accurately reflects true underlying skill. Finally, the data-joining process will be refined to reduce mismatches and missing data, allowing for a more complete and reliable dataset that strengthens overall model accuracy and validation efforts.

## Conclusion

Both pRating and pRating100 have clear, practical value for MLB coaching staffs, general managers, and scouts. Whereas most existing metrics isolate only specific aspects of a pitcher's performance, these ratings provide a single, holistic measure that captures command, efficiency, dominance, and overall effectiveness in one comparable value. Teams can use the metric diagnostically — for example, identifying pitchers with strong traditional stats but low pRating100 scores as potential regression candidates, or spotting those with underwhelming surface numbers but high pRating100 values who may be poised for a rebound. These kinds of performance and comparative analyses processes are crucially important in trades, transaction, contracts, and with the MLB Draft. The metric also enhances fan engagement by offering a single, intuitive indicator of pitching quality, rather than requiring interpretation of multiple, sometimes conflicting stats such as ERA, strikeouts, FIP, WAR, and wins. In short, pRating offers a unified and analytically meaningful tool with applications across front offices, coaching staffs, and fan communities throughout Major League Baseball.