Trust the Process

Using Process-Based Evaluation to Quantify Quarterback Decision-Making Quality in the Passing Game

---

Abstract

Quarterback decision-making remains one of the most scrutinized yet least quantifiable aspects of football performance. Existing efficiency metrics such as Expected Points Added (EPA) and Completion Percentage Over Expected (CPOE) capture outcomes but fail to isolate the quality of the decisions that precede those results. This study introduces a new framework, Expected Weighted EPA (xWEPA), designed to measure decision quality by combining play-by-play context, tracking-based situational factors, and quarterback-specific skill adjustments. Data from the 2021 through 2024 NFL seasons was analyzed to assess how xWEPA relates to team success, stability across games, and performance in high-leverage situations. The findings show that xWEPA identifies quarterbacks who consistently maximize decision quality beyond what traditional outcome-based metrics reveal. These insights demonstrate the potential for xWEPA to serve as a valuable tool for coaching evaluation, front office decision-making, and player development strategies.

1. Introduction

The development of sports analytics is often traced back to the Moneyball revolution in baseball, when the Oakland Athletics reframed player value around the process of run generation. That movement shifted attention from traditional statistics to measures that explained how players contributed to team success through repeatable, process-oriented actions. Since that time, baseball has evolved into a data-saturated environment in which technologies such as Statcast allow performance to be analyzed with extraordinary precision.

Football, by comparison, remains less developed in terms of process-based analytics. The most prominent efficiency measures, including Expected Points Added (EPA) and Completion Percentage Over Expected (CPOE), capture the results of plays rather than the quality of the decisions that produced those outcomes. This distinction is particularly critical at the quarterback position, which is the most scrutinized role in American sports. Quarterbacks are responsible for decisions that determine both the efficiency of individual plays and the success of entire offensive drives, yet existing statistics often fail to isolate decision-making from execution and surrounding context.

This study introduces Expected Weighted EPA (xWEPA), a framework designed to evaluate quarterback passing decisions through the integration of two components: the expected value of a throw as measured by expected EPA (xEPA) and the contextual difficulty of completing that throw as measured by quarterback-specific completion probability (qbxcp). By combining play-by-play outcomes with

situational features derived from Next Gen Stats tracking data, xWEPA establishes a scalable framework for assessing decision quality across quarterbacks, games, and seasons.

2. Methodology

2.1 Data Acquisition

Play-by-play data from the 2021 through 2024 NFL seasons was collected using the nflfastR package, which provides detailed information on play context and outcomes. To enhance the evaluation of quarterback decisions, these data were merged with Next Gen Stats tracking information available . The tracking data supplied additional situational variables, including route concepts, quarterback release timing, indicators of defensive pressure, air yards, and defensive coverage. The integration of play-by-play outcomes with tracking-based context allowed each passing attempt to be assessed not only by its result but also by the environment and decision surrounding the throw.

2.2 Play Filtering

The dataset was restricted to standard passing plays in order to maintain consistency in evaluating decision quality. Two-point conversions were excluded because of their atypical design and high-leverage nature. Plays occurring when the score differential exceeded 16 points were also removed. This restriction to “neutral game script” scenarios ensured that the analysis reflected decision-making under representative game conditions rather than desperation or garbage-time situations that often distort statistical relationships.

2.3 Core Variables

After filtering, the dataset retained variables that were essential for assessing quarterback efficiency and decision-making. These included play identifiers such as game, play, season, and quarterback identification. Contextual factors such as down, distance to first down, field position, and quarter were also preserved. Outcome measures included expected points added (EPA), win probability added (WPA), and the completion result of the pass. Passing details captured air yards, pass depth and direction, and baseline completion probability (cp). Advanced measures included pass-over-expected (pass_oe) and defensive coverage type. The inclusion of these variables provided the foundation for developing weighted efficiency metrics that incorporated both execution and decision context.

2.4 Adjusting for QB Skill (qbxCP)

Baseline completion probability, as reported by Next Gen Stats, was adjusted to account for quarterback-specific accuracy. Completion percentage over expected (CPOE) was calculated by stratifying attempts according to depth, location, and coverage type. A minimum attempt threshold was applied to ensure stability within each category, and missing values were conservatively imputed. These CPOE adjustments were applied in log-odds space using a logit transformation, which maintained probabilities within the valid 0 to 1 range. The resulting value, termed quarterback-specific completion probability (qbxcp), contextualized each passing attempt according to both situational difficulty and the demonstrated skill of the quarterback.

$$qbxcp = \sigma\left(\ln\frac{cp}{1-cp} + CPOE_{qb}\right)$$

2.5 Weighted Efficiency (WEPA)

While EPA captures realized play value and WPA reflects the impact on game outcomes, both remain heavily influenced by execution outside the quarterback's control, such as receiver performance or defensive breakdowns. To better isolate the quarterback's role, a weighted efficiency metric (WEPA) was introduced.

WEPA adjusts realized play value by incorporating the difficulty and risk profile of the passing attempt. Each throw's completion probability was combined with air yards, pass depth, and contextual win probability to weight its contribution to expected value. Plays where a quarterback attempted a difficult, high-value throw were credited proportionally more than plays where an easy, low-leverage pass was completed. Conversely, failed attempts on low-probability throws did not penalize quarterbacks as heavily, since the underlying decision carried positive expected value.

Formally, WEPA is defined as:

$$WEPA = EPA \times qbxcp$$

By anchoring realized play value to quarterback-specific probability of success, WEPA serves as an intermediate step between traditional outcome-based metrics and process-driven measures such as xWEPA

2.6 Modeling Expected Value (xWEPA)

Whereas WEPA anchors realized outcomes to quarterback-specific probability, xWEPA advances the framework by replacing realized play value with modeled expected play value. Expected EPA (xEPA) was modeled using predictors that included air yards, expected yards after the catch, and binary indicators for whether the play was expected to result in a first down, a field position change, or a touchdown. A linear regression model was cross-validated to predict EPA, and predictions were then combined with qbxcp to yield xWEPA as follows:

$$xWEPA = xEPA \times qbxcp$$

This formulation captures the expected value of a decision before execution unfolds, offering a process-oriented measure that isolates quarterback decision quality independent of outcome variance.

2.7 Stress Testing and Validation

The robustness of xWEPA was evaluated through several complementary analyses. At the game level, the quarterback with the higher per-attempt efficiency in EPA, WEPA, or xWEPA was labeled the "statistical winner" of the game, and these designations were compared with actual outcomes to test predictive alignment. At the team level, quarterback efficiency metrics were correlated with team scoring averages to assess the translation of individual decision quality into offensive production. To examine consistency, per-quarterback variance across games within a season was calculated for each metric, testing whether xWEPA reduced volatility relative to outcome-based measures. Finally, high-leverage situations were

analyzed by isolating the top ten percent of plays ranked by WPA and examining how xWEPA scaled with realized win probability impact in those moments.

2.8 Purpose

The methodology was designed to provide a multi-layered framework for evaluating quarterback performance. By integrating outcome data with situational context and quarterback-specific adjustments, the approach separates decision quality from raw results. This separation allows for a comprehensive assessment that spans play-level execution, game-level performance, and season-level consistency, offering insights into predictive power, reliability, and performance in critical situations.

3. Background

Quarterback evaluation has traditionally relied on outcome-based statistics. Standard box score measures such as yards, touchdowns, interceptions, and passer rating capture results but fail to reflect the quality of the decisions that produced them. The development of more advanced efficiency metrics, including Expected Points Added (EPA) and Win Probability Added (WPA), marked an improvement by incorporating game context and linking plays to their contribution toward winning. Completion Percentage Over Expected (CPOE) further advanced quarterback evaluation by framing accuracy in probabilistic terms, adjusting for situational difficulty.

Despite these advances, existing measures remain constrained by their retrospective nature. EPA and WPA are shaped by the realized outcomes of plays, which may be heavily influenced by factors outside the quarterback's control such as receiver execution, defensive breakdowns, or random variance. CPOE isolates accuracy but does not capture the broader decision of whether a pass should have been attempted in the first place. Each of these statistics provides value but ultimately remains oriented toward outputs rather than the process of decision-making.

Other sports have demonstrated the importance of shifting from outcomes to process. In baseball, the Statcast era emphasized granular measures such as exit velocity and launch angle, which evaluate the quality of contact independent of whether a batted ball results in a hit. This separation of skill from luck reshaped player evaluation by focusing on sustainable process-based indicators. Football, despite significant advances in player tracking data, has yet to establish an equivalent paradigm for the quarterback position.

The xWEPA framework aims to address this gap by quantifying the decision itself rather than its result. By integrating expected value with quarterback-specific completion probabilities, xWEPA moves quarterback evaluation toward a process-driven model. This approach complements existing metrics while providing a new lens through which to assess decision quality across plays, games, and seasons.

4. Results

The evaluation of xWEPA was conducted relative to established efficiency metrics (EPA per attempt and WEPA per attempt) to assess its predictive validity, stability, and performance in high-leverage contexts. Across all analyses, xWEPA demonstrated a meaningful balance between predictive strength and

statistical consistency, reinforcing its value as a process-oriented complement to traditional, outcome-based statistics.

4.1 Predictive Validity with Team Outcomes

The first analysis examined the relationship between quarterback efficiency metrics and team offensive production. Three scatterplots, one for each metric (EPA, WEPA, and xWEPA per attempt), illustrated their associations with team points per game. All three measures displayed positive relationships, confirming that quarterback efficiency contributes directly to scoring output. As anticipated, EPA and WEPA exhibited the strongest associations because they incorporate realized outcomes. However, xWEPA also maintained a significant positive relationship despite being independent of play results, indicating that decision quality itself translates into offensive productivity.

A complementary grouped bar chart compared win rates in head-to-head matchups between two qualified quarterbacks. For each matchup, the quarterback with the higher seasonal efficiency metric was designated the "statistical winner." Across the 2021 through 2024 seasons, the superior quarterback in EPA, WEPA, and xWEPA all achieved win rates greater than 58 percent. This finding underscores that all three measures, including the process-based xWEPA, reliably distinguish winning performance.

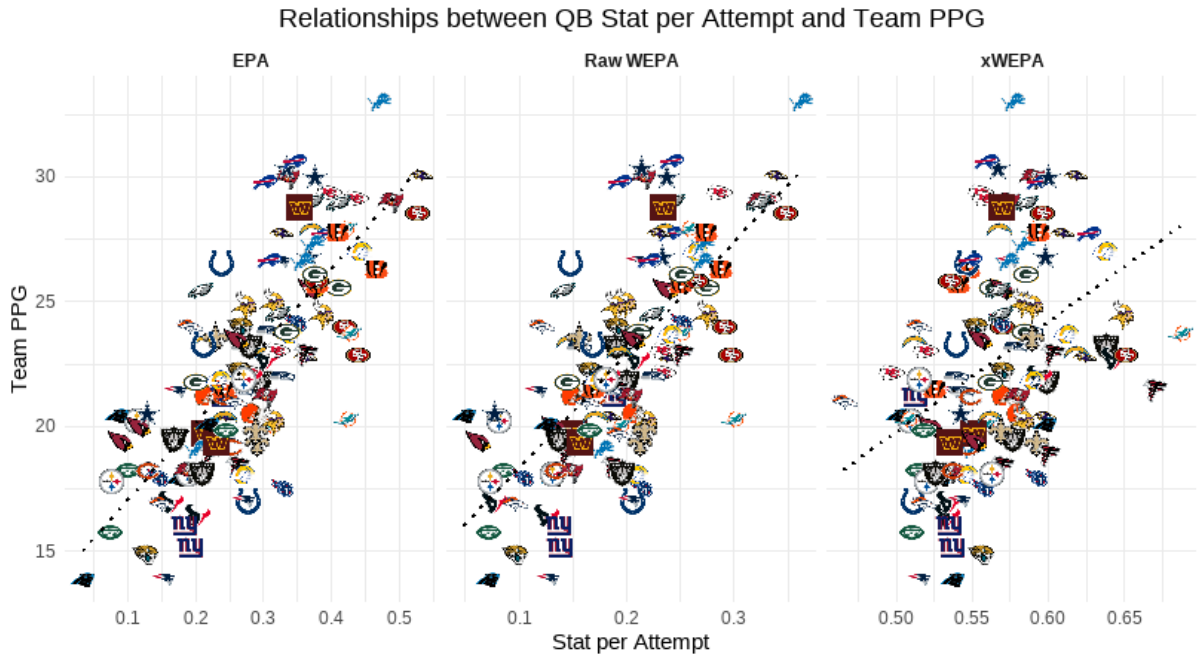

Figure 1. Scatterplots of EPA, WEPA, and xWEPA per attempt against team points per game. Each shows a positive trend, indicating that higher efficiency corresponds with greater offensive output.

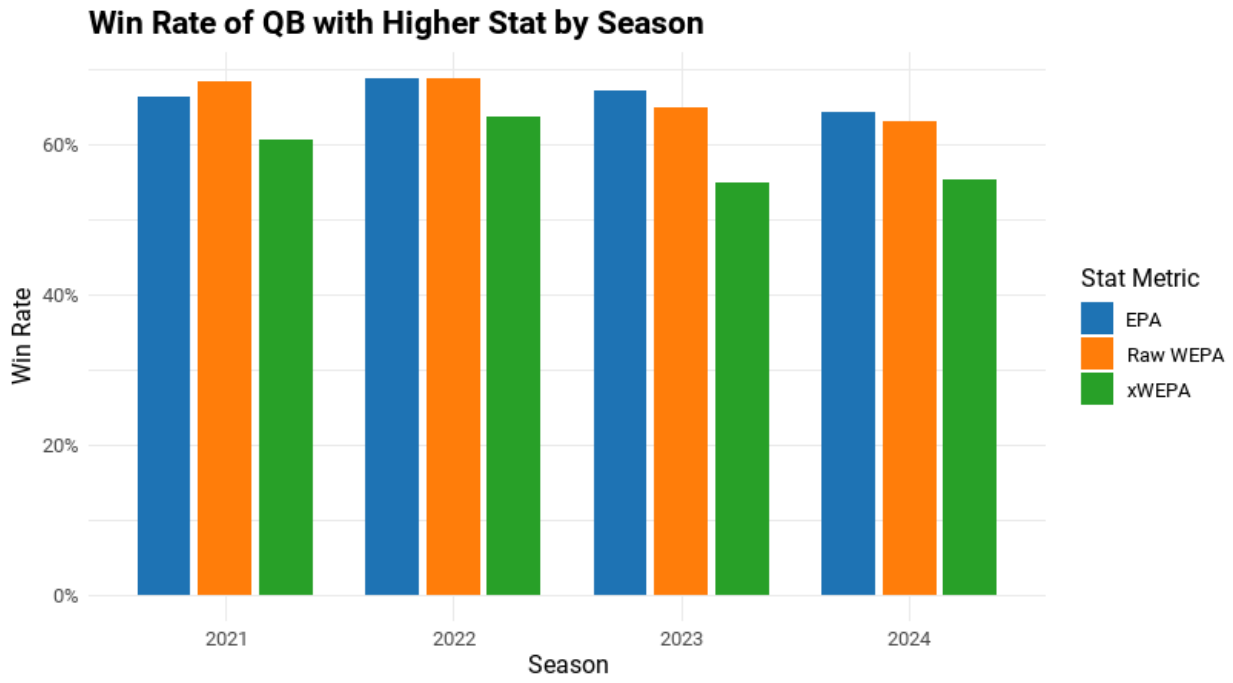

Figure 2. Grouped bar chart of win rates for quarterbacks with higher seasonal efficiency metrics in head-to-head matchups. All three measures exceed a 58 percent win rate from 2021 to 2024, highlighting their predictive relevance for game outcomes.

4.2 Variance and Consistency Across Games

A central aim of xWEPA is to reduce the influence of factors outside a quarterback's control. To test this, variance in game-to-game performance was compared across the three metrics using combined box and violin plots.

The results demonstrated a clear hierarchy of stability. EPA exhibited the greatest variance, averaging approximately 1.3 standard deviations from the mean across games. WEPA reduced this volatility, with variance closer to 0.75 standard deviations. xWEPA displayed the lowest variance, clustering tightly around 0.4 standard deviations.

These findings highlight the stability of xWEPA. By adjusting for expected outcomes and quarterback-specific skill, the metric filters out volatility introduced by receiver errors, exceptional yards after catch contributions, or defensive miscues. This improved consistency makes xWEPA a more reliable tool for longitudinal evaluation, a property valued in both academic and applied settings.

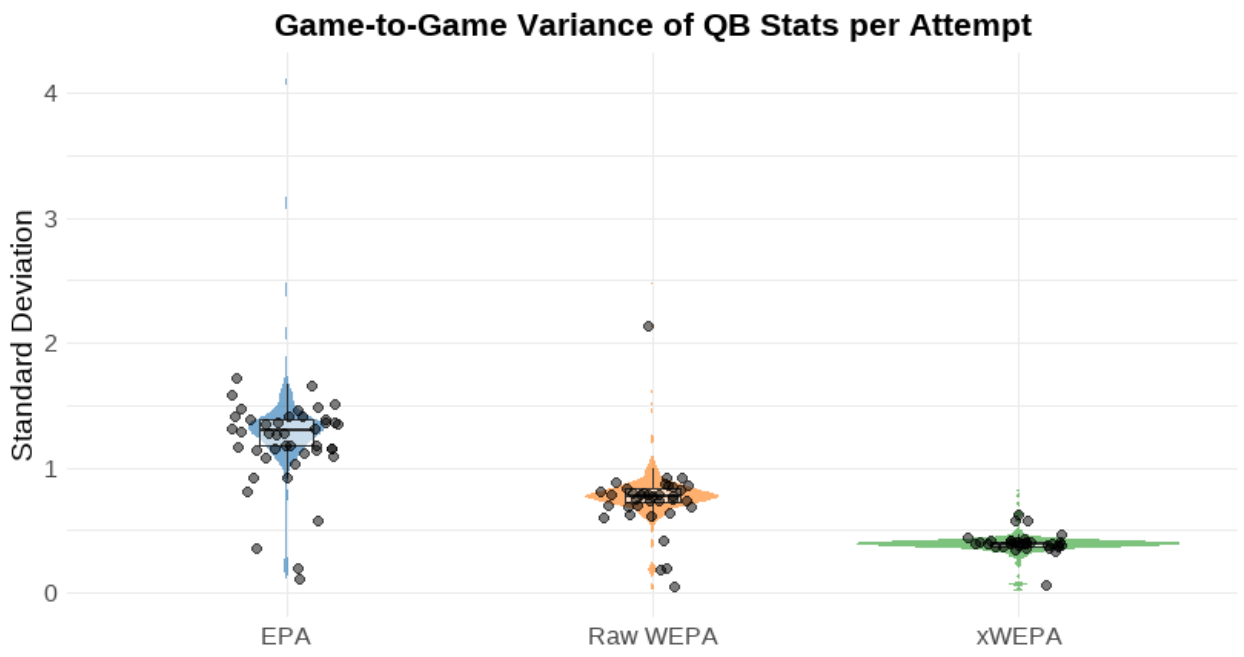

Figure 3. Box and violin plots comparing the variance of quarterback efficiency metrics across games. xWEPA shows substantially lower variance than EPA or WEPA, reflecting its greater stability and reliability.

4.3 High-Leverage Situations

The final analysis examined performance in high-leverage scenarios by isolating the top decile of plays ranked by win probability added (WPA). A scatterplot with a fitted trend line compared xWEPA with realized WPA in these critical situations. The relationship proved exceptionally strong, with an R^2 value of 0.935. This indicates that quarterbacks who consistently make high-quality decisions, as measured by xWEPA, are also the most effective at generating win probability when the stakes are highest.

This result demonstrates the practical value of a process-oriented framework. Whereas traditional metrics confirm what has already occurred, xWEPA identifies which quarterbacks are most likely to succeed in pivotal moments based on the quality of their decisions. In this way, xWEPA bridges the gap between statistical consistency and game-defining impact.

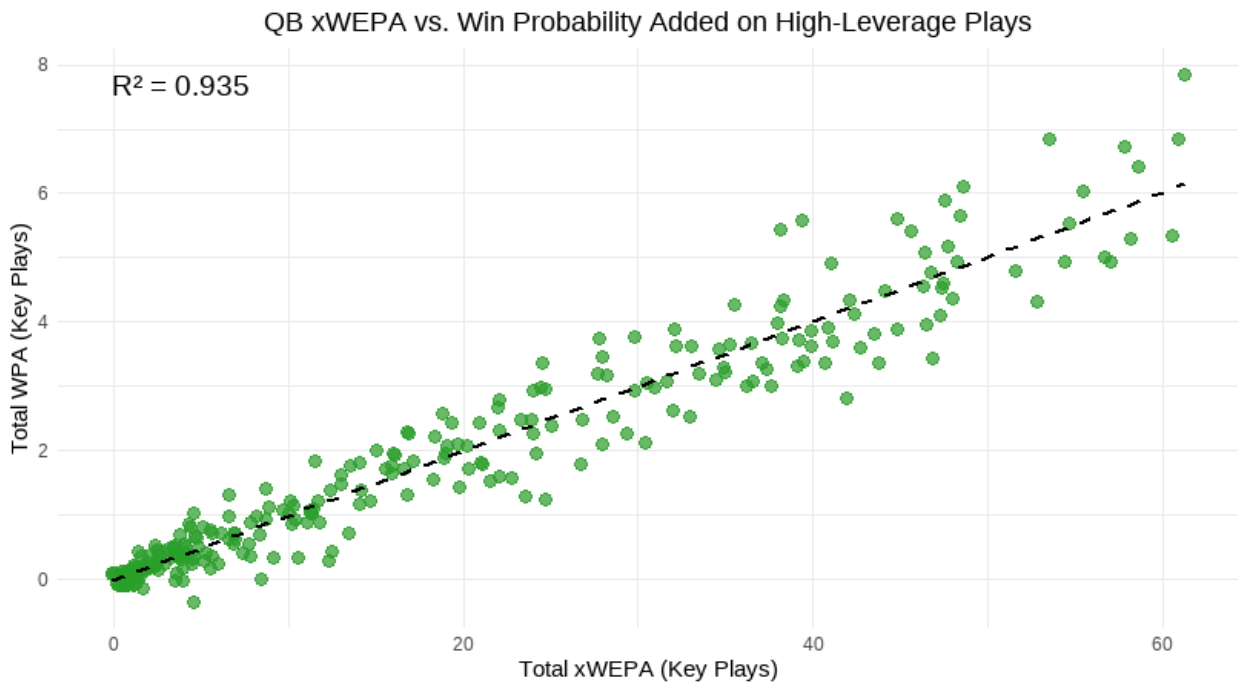

Figure 4. Scatterplot of xWEPA versus WPA in high-leverage situations. The strong relationship (R^2 = 0.935) demonstrates that quarterbacks with superior decision quality also maximize win probability when the stakes are highest.

4.4 Summary of Findings

The analyses support three principal conclusions. First, xWEPA retains meaningful positive associations with both team scoring and win percentage even though it excludes realized outcomes. Second, xWEPA substantially reduces variance relative to EPA and WEPA, making it a stable and consistent measure of decision quality. Third, xWEPA demonstrates a strong correlation with performance in high-leverage plays, validating its ability to capture the value of quarterbacks who consistently make sound decisions and deliver under pressure.

Together, these findings reinforce the guiding principle of the study: in the evaluation of quarterbacks, the process of decision-making provides critical insight that outcome-based metrics alone cannot reveal.

5. Quarterback Statistical Evaluation

With the framework established, the analysis turned to evaluating quarterbacks across multiple seasons using the developed metrics. All players from the 2021 through 2024 NFL seasons who attempted at least the qualifying minimum number of passes, defined by league standards as 14 attempts per team game (238 attempts per season), were included. This threshold ensured that only quarterbacks with a substantial sample of attempts were considered, allowing for reliable comparisons across players and seasons. Each qualified quarterback was evaluated in terms of EPA per attempt, WEPA per attempt, and xWEPA per attempt to capture both outcome-based efficiency and process-oriented decision quality. The table below presents the comparative results for all qualified quarterbacks from the 2024 season.

Consensus Rank	Quarterback	Team	Attempts	EPA/PA	Raw WEPA/PA	xWEPA/PA	EPA Rank	WEPA Rank	xWEPA Rank
1	L.Jackson		475	0.529	0.337	0.617	1	2	4
2	B.Purdy		405	0.437	0.295	0.651	4	6	1
3	J.Goff		480	0.469	0.361	0.576	3	1	12
4	B.Mayfield		494	0.486	0.322	0.573	2	3	13
4	S.Darnold		517	0.383	0.272	0.647	8	8	2
6	J.Hurts		404	0.409	0.321	0.585	6	4	10
7	J.Burrow		608	0.408	0.272	0.593	7	7	8
8	K.Cousins		408	0.360	0.250	0.602	10	10	5
9	J.Love		335	0.374	0.251	0.585	9	9	11
10	J.Herbert		471	0.305	0.192	0.618	15	19	3
10	J.Daniels		525	0.352	0.233	0.569	11	12	14
12	T.Tagovailoa		341	0.421	0.299	0.506	5	5	28
13	J.Allen		492	0.343	0.240	0.559	12	11	16
14	C.Stroud		547	0.278	0.181	0.593	17	20	7
15	D.Carr		255	0.286	0.215	0.569	16	15	15
16	G.Smith		536	0.325	0.225	0.539	13	14	20
17	P.Mahomes		617	0.319	0.233	0.519	14	13	24
18	M.Stafford		542	0.233	0.172	0.589	22	22	9
18	K.Murray		462	0.258	0.205	0.555	20	16	17
20	B.Young		275	0.216	0.124	0.598	23	26	6
20	D.Maye		249	0.267	0.201	0.540	19	17	19
22	R.Wilson		328	0.269	0.180	0.514	18	21	26
22	G.Minshew		279	0.208	0.195	0.525	24	18	23
24	A.Rodgers		523	0.243	0.164	0.517	21	23	25
25	D.Jones		303	0.182	0.135	0.534	26	24	22
26	C.Rush		259	0.127	0.077	0.542	27	28	18
27	C.Williams		451	0.127	0.133	0.536	28	25	21
28	B.Nix		502	0.190	0.121	0.514	25	27	27

Figure 5. Table displaying results of EPA, WEPA, and xWEPA per attempt for qualified quarterbacks in the 2024 NFL season.

5.1 Key Takeaways

Three principal observations emerge from the evaluation. First, quarterback rankings in EPA and WEPA tend to be similar because WEPA incorporates EPA in its calculation. Second, meaningful differences are often observed when comparing EPA or WEPA with xWEPA, which signals variation in decision quality relative to realized results. Third, individual performance in efficiency metrics may diverge from team outcomes due to external factors such as defensive strength, special teams contributions, rushing success, or schedule difficulty.

5.2 Interpretation

Although most quarterbacks maintain similar rankings between EPA per attempt and WEPA per attempt, many show clear discrepancies when xWEPA is considered. These differences illustrate the distinction between results and decision-making quality. Quarterbacks who rank significantly higher in xWEPA than in EPA or WEPA are often those whose outcomes were negatively affected by factors outside their control. Examples include receiver drops, fumbles, or tipped interceptions that suppress realized efficiency despite sound decision-making. Conversely, quarterbacks who rank much lower in xWEPA than in outcome-based metrics are often benefitting from external advantages. These may include significant yards after catch generated by receivers, successful completions on ill-advised/risky throws, or scheme-driven performance that inflates results beyond the quality of the underlying decisions.

A further layer of insight arises when comparing individual metrics to team success. In some cases, quarterbacks may grade strongly in decision quality while playing for teams that fail to convert those decisions into victories, often due to deficiencies in the skill position groups or on defense or special teams. In other cases, quarterbacks may appear high-quality while benefitting from a strong defense or favorable circumstances. For example, in 2024 the Denver Broncos achieved success primarily through defensive performance rather than offensive efficiency, while the San Francisco 49ers underperformed relative to the quality of their passing game led by Brock Purdy.

Together, these observations demonstrate the value of incorporating xWEPA into quarterback evaluation. The metric isolates decision quality from realized results, offering a more stable and fair assessment of quarterback performance across the league.

6. Applications

The utility of xWEPA spans several domains within football operations and analysis.

- **Coaching and Player Evaluation**
  The metric allows coaches to assess quarterback decision-making independent of outcome variance. Quarterbacks who consistently select throws with high expected value demonstrate sound judgment even when results fluctuate. This enables more precise film grading and targeted feedback focused on decision processes rather than purely on outcomes.
- **Front Office and Roster Construction**
  xWEPA supports draft scouting, free agent evaluations, and contract decisions by isolating a quarterback's contribution from receiver performance and random game variance. This capability is particularly useful for comparing quarterbacks across different systems or supporting casts, allowing personnel decisions to reflect skill rather than circumstance.
- **Game Strategy and Analytics Adoption**
  On a weekly basis, xWEPA identifies quarterbacks who maximize expected value, informing defensive game plans and offensive playcalling. The metric also translates complex decision analysis into an accessible tool for broadcasters and analysts, connecting technical evaluation with broader football discourse.

7. Conclusion and Future Work

This paper introduced xWEPA, a framework for quantifying quarterback decision-making in the passing game. By combining play-by-play outcomes with tracking-based context and quarterback-specific skill adjustments, xWEPA provides a process-oriented measure that complements traditional outcome-based efficiency metrics. Modeling results demonstrate that xWEPA retains predictive power with respect to win percentage and team scoring, reduces variance compared to EPA and WEPA by filtering out noise from skill player execution, and aligns strongly with performance in high-leverage plays, achieving an R^2 of 0.935 against WPA. These properties indicate that xWEPA effectively isolates decision quality, capturing consistency, situational reliability, and cognitive skill in quarterback play.

xWEPA can serve as a robust tool for coaches, front offices, and analysts seeking to evaluate quarterbacks beyond raw results. Future extensions include expanding route-level counterfactual modeling to assess the expected value of all available options, incorporating defensive strength adjustments to contextualize decision quality against opponent caliber, and extending the framework to mobile quarterbacks by integrating rushing decision-making alongside passing. By shifting the focus from outcomes to process, xWEPA represents a step toward more nuanced quarterback evaluation, emphasizing not only what occurred on the field but also how and why decisions were made.