Evaluating Playmakers at the 2026 World Cup:

Tailor-made versus off-the-shelf models of chance creation.

The opportunity-adjusted model of chance creation.

This article evaluates playmakers at the 2026 World Cup using the opportunity-adjusted model of chance creation already described in analyses of European Leagues[i] and the Women's Super League[ii]. We also take as given the rationale for using *chances created in open play* (CC) as the basic measure, in preference to assists or expected assist[iii]. However, the article is more than a simple *N+1* contribution that merely examines a different competition. Of course it is interesting to find new perspectives on the World Cup, but the more enduring contribution of this article is that we compare the usual model (which is tailor-made for each league or competition separately) against an off-the-shelf variant derived from an aggregate of different leagues. A portable version of the model would be a great labour saving as it requires only simple arithmetic to implement.

We use the same two-step procedure as in the previous articles. The variants differ only in how the first step is carried out.

(i) We first build a *model of the many*, which estimates the number of chances we would expect a typical player to create, given (a) his minutes on the pitch, (b) the crosses he gets to make, (c) the passes he gets to make in the final third, and (d) all other passes not in the final third. Each of (a) – (d) can be thought of as opportunities afforded a player to create chances. The model is nothing more than a set of regression parameters that weight each of (a) – (d) to give the best overall fit to the data, though we are using robust regression rather than OLS to capture typical performance unmasked from atypical performance, the precious outliers. (For those wanting the precise details, we use Tukey's biweight regression with $c = 4.685$, and hc3 correction for heteroskedastic errors, as in previous research.)

(ii) We then plug into the regression equation, for each player, his profile of minutes, crosses, *etc,* to calculate what would be typical of that profile. Finally, the playmaker is then judged by the size of Actual minus Typical, which is simply the regression residual (the CCexcess). If the residual is negative, he is underperforming compared with typical; but if the residual is positive, he is overperforming; and the larger that (positive) residual, the better the playmaker, we claim.

The tailor-made model.

In the **tailor-made version**, the regression parameters in step (i) are estimated from the 2026 World Cup data itself. But immediately we run into our first problem. Even the 2 finalists only played 8 games, whereas the 16 teams who didn't progress beyond the group stage only played 3 games. On average, the 48 teams played only 4.33 games (= 390 mins). In order to be able to include players from all teams, we set a filter that each player must have played for at least 225 mins in the competition; this is much lower than the filter of 450 mins used in our previous studies, but even so it left only N = 393 players in the analysis. So our first problem is that each person's data is based on very few minutes, and two more problems follow hard on its heels. Those few minutes are unevenly spread over the teams; and the quality of the teams varies greatly, meaning it should be easier for the major teams to create chances against minor nations, but be harder going in the opposite direction. There is therefore a lumpiness in quantity and quality. By contrast, leagues are based on more minutes (about ten times more, give or take), every team plays the same number of minutes, and the competition is relatively homogeneous. In conclusion, we would probably not be performing these analyses were it not the world's premier competition. Nonetheless, when we do, the model that estimates CC for each player is:

$$\mathbf{CCest} = -0.2353 + 0.004821\mathbf{Mins} + 0.04806\mathbf{Pass3} + 0.05539\mathbf{Crosses} - 0.01010\mathbf{PassX3} \quad (1)$$

where Pass3 are the number of passes he made in the final third, and PassX3 are the number of passes made elsewhere on the pitch. (All coefficients presented to 4 significant figures, and are statistically significant to at least the 5% level). Then the measure of playmaking ability is just the excess chances created:

$$\mathbf{CCexcess = CC - CCest,}$$

where CC is the number of chances that the player actually created.

The off-the-shelf model.

Given the potential problems in the tailor-made version of the model, we now present the **off-the-shelf** version. In previous research covering the Big 5 leagues and the English Football Leagues (to Tier 4), it was discovered that the all leagues gave similar regression weights. Furthermore, applying ('misapplying') other leagues' regression weights to, say, La Liga's data gave correlationally similar CCest values for each La Liga player. Over the entire matrix of 'misapplied' regression weights the median Pearson's R correlation = 0.9931. This extremely high level of agreement prompted the question of whether a single set of regression weights could be used with hardly any loss of accuracy. We now take this idea forward. Regression weights were estimated using the data from all the above leagues, stacked as if from a single mega-league, and using a filter of 225 mins to be consistent with the version just now tailor-made for the World Cup, which leaves N = 4314 players in the analysis. The regression equation thus derived is:

$$\mathbf{CCest} = 0.3740 + 0.003667\mathbf{Mins} + 0.04413\mathbf{Pass3} + 0.06872\mathbf{Crosses} - 0.009753\mathbf{PassX3} \quad (2)$$

Pearson's correlation between CCest (tailor made) and CCest (off-the-shelf) was R = 0.9971. Also, Pearson's correlation between CCexcess (tailor made) and CCecess (off-the-shelf) was R = 0.9947. We conclude that the tailor-made model we build using equation (1) is practically speaking the same as the off-the-shelf model built using equation (2).

In future research on new leagues and new seasons, we can simply plug numbers into equation (2) without the need for the prior step of building a regression model to determine what those regression coefficients should be. Nor is this off-the-shelf equation sensitive to the *minimum minutes* criterion. The next table shows the regression equations estimated using different filters. Again, no practical difference.

Filter			Regression coefficients				
(mins)	N	R^2	Const	Mins	Pass3	Crosses	PassX3
225	4314	0.818	0.3740	0.003667	0.04413	0.06872	-0.009754
450	3902	0.787	0.5468	0.003596	0.04445	0.06860	-0.009798
900	3116	0.734	0.9338	0.003434	0.04464	0.06945	-0.009772

Corresponding t-values, given the 225 filter, for Const, Mins, Pass3, Crosses, PassX3, are: 2.89, 18.93, 32.21, 14.73, –29.34, all significant; and similar t-values for the other filters. To give the model some context, if minutes alone are used to estimate CC, $R^2 = 0.351$, instead of 0.818 (filter = 225 mins). Deflating any kind of contributions to per 90 mins values is a standard way of adjusting for the opportunity of spending time on the pitch; but because it does not account for anything that actually happens when a player is on the pitch, it is clearly a limited adjustment.

Who excelled in chance creation at the World Cup?

Before the calculations reveal who had a good playmaking World Cup, some caveats are needed. First, Opta Analyst (where the raw data originate) does not include the third place playoff game in its Passing / Chance Creation data, a strange omission, particularly since they do include it for Attacking / Goals / Shots data. The analyses are therefore for 103 of the 104 games. Second, although the tailor-made model converges with the off-the-shelf model, that convergence is probably due to there being a sufficient number of players in the analysis rather than that any given player's CC is a precise measure of his individual performance. This is a very important distinction. It means that the *model of the many* tells us accurately how CC varies with minutes, passes, *etc.* as a general trend. But any given player's CCexcess is less precisely measured in the World Cup than in the leagues because of fewer matches and minutes. In other words, if we want to know whether a player outperforms the typical, it would be better to base that evaluation on 3000 minutes of play rather than 300 minutes. On the other hand, the *model of the many* (applying the 225 minutes filter) is based on 137,182 player-minutes of game time; and 19,114 chances created. That is plenty enough to discern the general rules for how CCs come into being. Following our own advice, we use the off-the-shelf version of the model to calculate CCest, and the top 30 playmakers by CCexcess are shown in the next table.

World Cup 2026					
	Mins	CC	CCest	CCexcess	CCexcess in IQRS
Leandro Trossard	510	17	6.58	10.42	4.89
Désiré Doué	353	13	4.80	8.20	3.87
Kylian Mbappé	608	15	7.98	7.02	3.33
Mohamed Salah	428	12	5.13	6.87	3.26
Sadio Mané	364	12	5.22	6.78	3.22
Rayan	289	10	3.37	6.63	3.15
Julián Quiñones	414	10	3.92	6.08	2.90
Pau Cubarsí	750	8	2.25	5.75	2.75
Yan Diomande	331	10	4.31	5.69	2.72
Chris Wood	270	7	1.89	5.11	2.45
Ousmane Dembélé	552	14	9.16	4.84	2.33
Noni Madueke	288	9	4.36	4.64	2.24
Lucas Paquetá	236	8	3.48	4.52	2.18
Hans Vanaken	318	7	2.51	4.49	2.17
Breel Embolo	506	8	3.52	4.48	2.17
Bruno Guimarães	419	9	4.61	4.39	2.12
Jhon Arias	381	8	3.69	4.31	2.09
Teboho Mokoena	270	6	1.70	4.30	2.08
Dani Olmo	495	10	5.71	4.29	2.08
Viktor Gyökeres	360	8	3.73	4.27	2.07
James Rodríguez	317	9	5.19	3.81	1.86
Davinson Sánchez	480	5	1.24	3.76	1.84
Rodri	726	11	7.36	3.64	1.78
Pedro Vite	360	8	4.47	3.53	1.73
Luis Díaz	478	9	5.56	3.44	1.69
Brian Brobbey	225	5	1.59	3.41	1.67
Alexander Isak	358	7	3.60	3.40	1.67
Vinícius Júnior	441	9	5.62	3.38	1.66
Brahim Díaz	462	8	4.70	3.30	1.62
Mostafa Zico	375	6	2.71	3.29	1.62

The last column is CCexcess re-expressed in terms of Inter Quartile Range units above (+) or below (–) the median. Lighter green highlighting corresponds to players whose CCexcess >= 2.0 IQRs (termed *outliers*). Dark green highlighting are players whose CCexcess >= 3.5 IQRs above median (termed *far outliers*).

Illustration: Four players compared in detail.

Few would have expected Trossard to top the list, but as seen through the lens of the opportunity-adjusted model, he is a worthy winner. However, a notable absence from the list is Messi, particularly considering his unadjusted CC of 15 is second equal with Mbappe for the tournament. The Argentinian team worked tirelessly for Messi to give him opportunities to influence the game; and they succeeded because his CCest (11.73) was the highest of anyone in the tournament. Messi himself added his own value, but only by about 3 CCs. Meanwhile, the Belgian team did not similarly privilege Trossard (CCest = 6.58), so that his final tally of 17 CCs is a huge excess. But to fully understand how Trossard and Messi compare, and what really counts in the model, it is helpful to follow through the calculations by hand, which we do for four players, below.

Player	CC	Const	Mins	Pass3	Crosses	PassX3
Leandro Trossard	17	1	510	108	7	93
Lionel Messi	15	1	740	198	20	151
Pau Cubarsí	8	1	750	109	1	590
Lamine Yamal	5	1	615	155	24	112
	CCest					
Leandro Trossard	6.58	0.37	1.87	4.77	0.48	-0.91
Lionel Messi	11.73	0.37	2.71	8.74	1.37	-1.47
Pau Cubarsí	2.25	0.37	2.75	4.81	0.07	-5.75
Lamine Yamal	10.03	0.37	2.26	6.84	1.65	-1.09
Regression coefficients:		*0.3740*	*0.0037*	*0.0441*	*0.0687*	*-0.0098*

The top half of the table is the raw data, while the bottom half is that data multiplied by the regression coefficients in the bottom line. Rows are then summed to give CCest. From looking at the bottom half, we see that it is Pass3 that most distinguishes between Trossard and Messi in terms of their expected chances created. From the regression coefficients, we can think of crossing as a special kind of passing in the final third, but with a slightly (≈ 50%) higher tariff (0.0687 / 0.0441 ≈ 1.5). However, crosses occur very much less frequently, so their overall impact is diminished. Hence the great disparity in the number of crosses these players make is not followed though into a great disparity in impact on CCest.

Yamal's profile is very similar to Messi's, resulting in a similar CCest. However, Messi created 15 chances, but Yamal only 5, resulting in a *negative* CCexcess (*i.e.*, a deficit). It's great to be given the ball in attacking areas, because the headline figures of goals, assists, and chances created will surely increase. This can lead to a player seeing even more of the ball, scoring even more goals, and so on, in a positive feedback loop. But here is the harsh truth: if opportunities are not used, that player could have been replaced with a player performing at average, with even better results. Furthermore, the more opportunities given, the larger that deficit can be. In Yamal's defence, at the start of the competition he was returning from injury; he posted a generous CCexcess in La Liga last season (bordering on being a positive outlier – in IQR units CCexcess = 1.956); and few games are played at the World Cup meaning the random fluctuations loom larger.

Happily, there is also a *vice versa*. For instance, Cubarsi's eight CCs would probably be lost in the crowd, but when adjusted for the limited opportunities afforded a centre-back, he was in the top 10 performers on this attribute alone. Add this to the other components of his game, and we can see why his market value jumped by €20m this summer, according to Transfermarkt.

Conclusions.

We have built two variants of the opportunity-adjusted model of chance creation differing at step (i). In the first, tailor-made version, we use the data from the World Cup itself to determine the numbers of chances that each player in the tournament might be expected to create. In the second, off-the-shelf version, we use data from 9 European leagues to make that estimation. The two models are highly convergent. This shows that a small amount of data about a lot of players can be used to create a stable *model of the many* (as in the tailor-made version). But even more importantly, it means that we no longer need that lengthy, knowledge-intensive step (i), but can replace it with a ready-made equation that only involves multiplying a few numbers and adding the results, as carried out in the Trossard / Messi / Cubarsí / Yamal table. Choosing the 450mins filter and printed in red so it cannot be missed, that equation is:

CCest = 0.5468 + 0.003596**Mins** + 0.04445**Pass3** + 0.06860**Crosses** – 0.009798**PassX3**

This single-equation version of step (i) makes the model entirely more user-friendly.

However, when it comes to assessing each player individually against the *model of the many*, we need to be more cautious, because each person plays in so few matches in the World Cup, and the quality of opposition is so uneven. Despite these caveats, we can safely conclude that Trossard had a better tournament than Yamal in terms of chance creation. But what will be remembered in years to come is the medal around Yamal's neck.

John Doyle
Football Analytica
12/09/2026.

---

[i] Evaluating Playmakers in 9 Leagues.
[ii] Evaluating Playmakers in the Women's Super League (WSL).
[iii] Use Chances Created, not Assists or Expected Assists.