Universität Münster

Institut für Organisationsökonomik

Diskussionspapier des
Instituts für Organisationsökonomik

3/2026

Individual Skill or Just Luck?

Decomposing Individual Finishing and Shot-Stopping Performance in European Football

Tom Böttger/Lars Vischer

Discussion Paper of the
Institute for Organisational Economics

Diskussionspapier des
Instituts für Organisationsökonomik
3/2026

März 2026

ISSN 2750-4476

Individual Skill or Just Luck?

Decomposing Individual Finishing and Shot-Stopping Performance in European Football

Tom Böttger/Lars Vischer

Abstract

Expected goals (xG) are widely used to evaluate football performance. However, it is unclear whether players who consistently outperform their xG do so because of skill or because of luck. To our knowledge, this paper provides the first meta-analytic estimate of between-player heterogeneity in finishing and shot-stopping overperformance, extending existing team-level skill-luck decompositions provided by Holzmeister & Johannesson (2025) to the individual player level. We analyse data from the English Premier League, La Liga, Bundesliga, Serie A, and Ligue 1 across eight seasons (2017/18–2024/25), totalling 3,202 outfield player-seasons and 932 goalkeeper-seasons. Using restricted maximum likelihood (REML) random-effects meta-analysis, we find that between-player heterogeneity accounts for 6.753–29.945% of total variance in finishing overperformance (average $I^2 \approx 21\%$) and 7.690–34.344% in shot-stopping (average $I^2 \approx 25\%$). Skill is detectable but modest. A complementary persistence regression shows that part of this overperformance carries over to the next season, both for outfield players ($\beta = 0.110$, $p < 0.01$) and goalkeepers ($\beta = 0.185$, $p < 0.01$). From these coefficients we derive that about 52% of a player's finishing skill signal is retained from one season to the next. For goalkeepers, the corresponding figure is roughly 74%. A single season of xG overperformance is too noisy to serve as a reliable indicator of player ability. Identifying genuine skill requires observations over multiple seasons.

JEL Codes: C12, C23, J24, L83, Z20, Z21, Z22

Keywords: Expected Goals, Football, Meta-Analysis, Performance, Skill

Individuelles Können oder einfach nur Glück?

Zerlegung individueller Torabschluss- und Torwartleistung im europäischen Fußball

Zusammenfassung

Erwartete Tore (xG) sind mittlerweile ein zentraler Maßstab zur Bewertung von Leistung im Fußball. Ob Spieler, die ihre xG-Werte regelmäßig übertreffen, dies aufgrund von Können oder Zufall tun, ist jedoch ungeklärt. Diese Arbeit liefert die erste metaanalytische Schätzung der Heterogenität zwischen Spielern im Torabschluss und in der Torwartleistung und erweitert damit die Dekomposition von Können und Zufall auf Teamebene von Holzmeister & Johannesson (2025) auf die individuelle Spielerebene. Grundlage sind Daten aus der English Premier League, La Liga, Bundesliga, Serie A und Ligue 1 über acht Spielzeiten (2017/18–2024/25) mit insgesamt 3.202 Feldspieler-Saisons und 932 Torwart-Saisons. Mithilfe einer REML-Random-Effects-Meta-Analyse zeigen wir, dass die Heterogenität zwischen Spielern 6,75–29,95 % der Gesamtvarianz in der Abschluss-Überperformance erklärt (durchschnittliches $I^2 \approx 21$ %) und 7,69–34,34 % bei Torhütern (durchschnittliches $I^2 \approx 25$ %). Können ist nachweisbar, fällt aber gering aus. Eine ergänzende Regressionsanalyse zur Persistenz zeigt, dass ein Teil der Überperformance in die Folgesaison übertragen wird. Das gilt sowohl für Feldspieler ($\beta = 0{,}110$, $p < 0{,}01$) als auch für Torhüter ($\beta = 0{,}185$, $p < 0{,}01$). Daraus lässt sich ableiten, dass etwa 52 % des Abschluss-Signals von einer Saison zur nächsten erhalten bleiben. Bei Torhütern sind es rund 74 %. Eine einzelne Saison xG-Überperformance ist damit zu verrauscht, um individuelles Können verlässlich abzubilden. Eine belastbare Einschätzung erfordert Beobachtungen über mehrere Spielzeiten.

Im Internet unter:

http://www.wiwi.uni-muenster.de/io/forschen/downloads/DP-IO_03_2026

DOI: 10.17879/60948663147

Universität Münster
Institut für Organisationsökonomik
Scharnhorststraße 100
D-48151 Münster

Tel: +49-251/83-24303 (Sekretariat)
E-Mail: io@uni-muenster.de
Internet: www.wiwi.uni-muenster.de/io

Individual Skill or Just Luck?

Decomposing Individual Finishing and Shot-Stopping Performance in European Football

1. Introduction

Expected goals (xG) has become one of the most important metrics in the evaluation of football performance in the last decade. The growing importance of the xG metric is not only apparent in its increasing discussion and media coverage, but also in the academic literature (Anzer & Bauer 2021, Holzmeister & Johannesson 2025, Rathke 2017, Robberechts & Davis 2020). When the level of performance of an individual player consistently differs from what is statistically anticipated, the explanation is often unclear. Does consistent overperformance indicate an extraordinary finishing skill or is it a statistical artefact of random variation? Due to the low-scoring nature of football, goal-scoring events have a great impact on the final result which distinguishes football from higher-scoring sports where the law of large numbers links performance more closely to the underlying skill of teams and players (Heuer et al. 2010, Mauboussin 2012). Observed overperformance thus blurs the distinction between a potential skill signal and substantial random noise or luck, which creates a fundamental identification problem.

Holzmeister & Johannesson (2025) recently examined this problem at the level of teams, using random-effects meta-analysis to decompose deviations from expected performance into skill and luck. Their results indicate that approximately 29–38% of the variation in goals overperformance of a team is attributable to skill, which is described as systematic differences between teams. The authors suggest extending their methodology to the less aggregated level of individual player performance for future research. Building on this suggestion, this paper aims to answer the following research question:

To what extent do individual players exhibit systematic skill in finishing and shot-stopping, as measured by overperformance relative to expected goals?

We extend the research from the aggregated team level to the individual player level for both outfield finishing performance and goalkeeper shot-stopping performance. Importantly, our analysis focuses exclusively on finishing quality conditional on a shot being taken. We do not aim to capture broader dimensions of offensive player performance, such as the ability to create shooting opportunities, positional play, or assist quality. These aspects represent distinct skills that are not reflected in xG-based overperformance metrics. We use a restricted maximum

likelihood (REML) estimation within a random-effects meta-analysis. We decompose the variance in players' overperformance into between-player heterogeneity and within-player variance. In order to investigate whether skill persists between seasons, we conduct a persistence analysis as a complementary test, following the logic of Miles et al. (2013) and Getty et al. (2018).

The analysis is based on player-level data from Europe's five leagues from the years 2017/18 to 2024/25, spanning 8 seasons and 3,202 outfield player-seasons and 932 goalkeeper-seasons. The results indicate that between-player heterogeneity explains approximately 7–30% of total variance in finishing overperformance and 7–35% for shot-stopping overperformance. Thus, there are systematic differences between players, but the differences are modest in magnitude. In the persistence analysis, a positive relationship between consecutive seasons of overperformance is noted for outfield players ($\beta = 0.110$, $p < 0.01$) and goalkeepers ($\beta = 0.185$, $p < 0.01$), indicating that overperformance is partly consistent across seasons due to a skill component of players. To our knowledge, this paper is the first to conduct a systematic player-level decomposition of skill and luck for xG-based performance metrics using meta-analytic techniques.

2. Theoretical Background

Distinguishing between skill and luck in competitive outcomes has long been studied in economics, psychology, and sports science. Mauboussin (2012) argues that most activities combine elements of both, and that the relative contribution of skill and luck can be estimated empirically. In sport, observed performance can therefore be understood as a combination of systematic (skill) and stochastic (luck) factors. For a distinction between the two, there is a need for structural decomposition or a high number of observations.

Miles et al. (2013) attempt to formalise this logic by proposing four tests for identifying skill and luck in competitive environments. One of them involves temporal persistence. If outcomes are serially correlated across periods, it is interpreted as evidence for skill. Getty et al. (2018) build on this logic by analysing different leagues of professional sports (MLB, NBA, NFL, NHL), fantasy sports, and mutual fund returns. They position all activities along a skill-luck continuum. They observe systematic differences in the performance correlation across time in each sport. According to the analysis, basketball is the most skilful, and hockey is the most luck-based activity.

In the context of football, there are several lines of research regarding skill and luck. One examines the fundamental stochastic nature of scoring a goal. Heuer et al. (2010) show that goals in professional football are comparable to a Poisson process. Goal scoring rates are affected by the strength of the team, suggesting that match-to-match variations in goal scoring are largely random. According to Wunderlich et al. (2021), almost half of the goals scored in the English Premier League show a high degree of randomness, e.g. deflections, goalkeeping mistakes, and lucky rebounds. Lames (2018) similarly emphasises the empirical importance of chance involvement in goal-scoring. These findings collectively establish that individual-level finishing outcomes contain a large stochastic component, motivating the use of formal decomposition methods to quantify how much skill is involved.

The second strand focuses on the development, application and accuracy of expected goals (xG) models. Anzer & Bauer (2021) present an xG model that integrates synchronised tracking and event data from the German Bundesliga, outperforming all previously published xG models in predictive accuracy. Robberechts & Davis (2020) investigate the relationship between data availability and the ability to train good xG models. They examine the effects of training set size, data recency, league specificity, and data source on model performance. The xG model can be used to measure actual performance by calculating the difference between realised goals and expected goals. This represents the overperformance metric used by Holzmeister & Johannesson (2025) to measure skill and luck at the team level. As we extend their approach to the individual player level, overperformance relative to expected goals forms the basis of our analysis.

For goalkeepers, instead of xG a similar metric called post-shot expected goals (PSxG) is used. De-la-Cruz-Torres et al. (2025) propose an alternative model which they call expected goals on target (xGOT). Their model includes shot placement, trajectory, but also goalkeeper positioning to estimate save difficulty, thereby separating shot quality from factors a goalkeeper can influence (De-la-Cruz-Torres et al. 2025). Assuming that player ability is involved in goalkeeping provides the methodological basis for evaluating individual shot-stopping ability relative to a model-implied benchmark, analogous to the use of xG for outfield finishing.

In football, the individual skill of players is accumulated to a teams' performance, where skill and luck can be also noted. Brechot & Flepp (2020) evaluate team performance in European club football using performance data compared to xG metrics. They state that xG-based measures predict future match results significantly better than raw outcome-based measures

implying that expected performance captures a meaningful amount of a team's systematic quality.

Holzmeister & Johannesson (2025) represent the most directly relevant predecessor study. They use random-effects meta-analysis across three seasons in seven European leagues to break down deviations from expected performance at the team level into components of skill and luck. Their I^2, which represents the estimated share of skill, ranges from approximately 29–55% for overperformance measured in points and 29–38% for goals overperformance. That means that roughly one-third to one-half of the team-level variation in overperformance is attributable to systematic differences between clubs and therefore skill. For more granular submetrics, offensive overperformance (goals scored minus xG), they estimate I^2 values of approximately 22–36%. For defensive overperformance (xG conceded minus goals conceded) the I^2 value is about 20–31%. These values give natural benchmarks for the outfield and goalkeeper estimates at the player level in our research. They conclude that luck plays a quantitatively important role even at the team level and show the impact of luck on final season rankings by simulating its contribution across 10,000 seasons per league and reporting the possible shifts in team positions. In contrast, Sarkar & Kamath (2023) compare actual and expected performance in Europe's five major leagues and find that luck does not systematically influence final league standings. Crucially, Holzmeister & Johannesson (2025) suggest extending their approach to individual players as a direction for future research, which we address with this study.

Baron et al. (2024) address finishing skills on an individual level by exploring off-target shots. They explain how a player's shooting ability can be appreciated through unsuccessful shots which miss the target, in addition to shots that lead to a goal or save. This implies that finishing skill is a real, measurable attribute of individual players. Davis & Robberechts (2024) provide a complementary perspective by analysing biases in xG models themselves. They identify three sources of concern: limited sample sizes (only about 9% of players accumulate more than 50 shots per season), the xG metric needs to assess a huge variety of shot types, and interdependencies in the training data that systematically compress the measured overperformance of top finishers. Their research emphasises the need to contain possible confounding factors in the skill-luck decomposition of performance.

An alternative idea to evaluate individual player ability is suggested by Kharrat et al. (2020), who adapt the plus-minus ratings approach from basketball and ice hockey for football. They propose two different metrics: an expected goals plus-minus rating (xGPM), which identifies players contributing to a team's scoring chances, and an expected points plus-minus rating

(xPPM), which captures a player's contributions to match outcomes. They use data from eleven European leagues and find that their ratings correlate at 0.35 from one season to the next, providing evidence of moderate temporal persistence in individual player performance. While this approach captures a player's broader contribution to team performance, it does not allow for a direct decomposition of overperformance relative to a shot-quality benchmark, which is the focus of the present study.

Analysing the balance between skill and luck also carries significant economic implications that extend well beyond sports analytics. Gauriot & Page (2019) provide direct empirical evidence of this distortion in professional football. They show that random outcomes, in their case shots hitting the post, causally affect player evaluations: players who score on post shots receive significantly more playing time and higher expert ratings than others. This demonstrates that evaluators systematically overweight realised outcomes without distinguishing sufficiently between signal and noise. This systematic confusion between signal and noise further underscores the need for precise empirical decomposition at the individual player level.

3. Hypotheses

Drawing on this literature, we develop four hypotheses to guide the empirical analysis. The first hypothesis attempts to answer whether outfield players systematically differ from each other when it comes to finishing chances. If finishing overperformance was purely attributable to luck, all between-player heterogeneity would be zero. This provides evidence of individual finishing skill (Baron et al. 2024) and observes skill on team-level (Holzmeister & Johannesson 2025). We therefore assume that there is heterogeneity at the player level.

H1: Individual outfield players differ systematically in their finishing ability, such that overperformance relative to expected goals is not purely attributable to random variation.

The second hypothesis applies the same logic to goalkeepers. Shot-stopping involves both physical and cognitive skills. De-la-Cruz-Torres et al. (2025) demonstrate that xGOT models can reliably capture differences in goalkeeper performance, lending empirical support to PSxG-based metrics as a valid measurement approach for shot-stopping ability. If goalkeeper overperformance is not completely random, systematic between-goalkeeper differences should be observable.

H2: Individual goalkeepers differ systematically in their shot-stopping ability, such that over-performance relative to post-shot expected goals is not purely attributable to random variation.

The third hypothesis compares the two positions. Outfield finishing is characterised by high situational variability such as shot types, distances, body parts, and game contexts. In addition, the per-season sample sizes can be relatively low for many players, both of which complicate the isolation of individual skill (Baron et al. 2024; Davis & Robberechts 2024). Further supporting this, Wunderlich et al. (2021) find randomness in almost half of all goals scored in the Premier League. Goalkeepers, by contrast, face a more uniform task. Their performance is therefore assessed over a more homogeneous set of situations. In contrast to outfield players, a goalkeeper faces every shot on goal which yields a larger number of relevant events per season and therefore more precise individual estimates.

H3: Goalkeepers exhibit a more pronounced skill component in shot-stopping than outfield players exhibit in finishing.

The fourth hypothesis tests a complementary indicator of skill. Following Miles et al. (2013) and Getty et al. (2018) on the temporal persistence of skill, we argue that if overperformance contains a genuine skill component, a player's overperformance in one season should positively predict their overperformance in the following season. Under the null hypothesis of pure randomness, consecutive season overperformance for the same player would be entirely uncorrelated.

H4: Individual finishing and shot-stopping overperformance persists across seasons, indicating the presence of a stable, skill-based component in player performance.

4. Data and Methodology

Our empirical strategy follows Holzmeister & Johannesson (2025), who apply random-effects meta-analysis at the team level. The methodology has its roots in the biostatistics and social sciences. Borenstein et al. (2010) provide a foundation for the field and random-effects meta-analysis, making distinctions between fixed-effect and random-effects frameworks and specifying when each is appropriate. Viechtbauer (2005) considers the random-effects model and various alternative variance estimators and shows that for a given estimator, the REML estimator is a favourable combination in terms of being approximately unbiased and efficient. Higgins & Thompson (2002) define I^2 as the pro-portion of total variation in effect estimates due to

genuine heterogeneity rather than sampling error. I^2 is therefore a scale-independent measure that is comparable across different meta-analyses.

For our analysis, we collected player-level performance data from five of the largest European football leagues: the English Premier League, Spanish La Liga, German Bundesliga, Italian Serie A, and French Ligue 1, covering the period from the 2017/18 season to the 2024/25 season. Player-level data were obtained from FBref using web-scraping through Python. Data were collected prior to January 2026. At that point, Stats Perform terminated its data-sharing agreement with FBref, as a result, Opta-based advanced metrics like xG values are no longer publicly available on the platform. The underlying xG model estimates goal-scoring probability based on shot characteristics including location, body part, type of preceding pass, and type of attack and does not consider the identity or the quality of the player. The variables we use for outfield players are npxG (non-penalty expected goals) and for goalkeepers, post-shot expected goals (PSxG) at the individual match level.

To measure the finishing overperformance for outfield players, we define finishing overperformance as OP = NPG − npxG for each game. NPG denotes non-penalty goals scored. Following Holzmeister & Johannesson (2025), we use the term 'overperformance' as a shorthand for deviations from expected performance in both directions. OP can take positive values (outperformance) as well as negative values (underperformance). OP captures whether a player scores more (positive values) or fewer (negative values) goals than would be expected based on the quality of their shooting opportunities. For the analysis, we do not consider penalty kicks for outfield players as they are different from other finishing situations, since they involve a different type of decision-making.

For goalkeepers, we define shot-stopping overperformance as GK_OP = PSxG − GA per game, where PSxG is expected goals conceded, and GA is actual goals conceded. If GK_OP is positive, then the goalkeeper conceded fewer goals than expected based on the quality of the shots faced, hence demonstrating good shot-stopping. If GK_OP is negative, then the goalkeeper is considered to underperform.

We apply certain cutoff points at the player-season level to exclude observations with insufficient sample size. For outfield players, a player-season is included if a player participated in at least 10 games, with a minimum playing time of 30 minutes each, and a non-penalty expected goal (npxG) of at least three during the entire season. In practice, this threshold filters out defenders and deep-lying midfielders, who rarely accumulate substantial non-penalty expected

goals. For goalkeepers, the thresholds are at least 10 games with a minimum playing time of 45 minutes and a cumulative PSxG of at least three.

For each player-season, we derive mean_OP and the standard error by using the formula $\text{se_OP} = \text{sd_OP} / \sqrt{n}$, where n is the number of games eligible. The formula assumes independence across games within a player-season, which is approximately satisfied if goals follow a Poisson process conditional on shot quality (Heuer et al. 2010). This assumption may be violated for players experiencing sustained hot or cold streaks, or when a tactical change alters shot quality over the course of a season. If a player appears in two leagues in one season we create two player-season observations. The same procedure is applied to goalkeeper-seasons.

Following Holzmeister and Johannesson (2025), we apply their meta-analytic method at the player level instead of the team level. The meta-analytic approach treats each eligible player-season as an independent study. All meta-analyses are conducted using Stata 18 with restricted maximum likelihood (REML) estimation, which provides an approximately unbiased estimator with favourable efficiency properties (Viechtbauer 2005).

Formally, the observed mean overperformance for player *i* in season *s* is modelled as:

$$y_{is} = \mu_s + u_i + e_{is} \quad (1)$$

where μ_s is the pooled mean overperformance in season *s*, u_i captures between-player heterogeneity (assumed normally distributed with variance ${\tau_s}^2$), and e_{is} represents within-player sampling error (with player-season-specific variance ${\sigma_{is}}^2$). The within-study variance is estimated from the squared standard error of overperformance for each player-season. The total variance of overperformance is defined as:

$$v_{is} = {\sigma_{is}}^2 + {\tau_s}^2 \quad (2)$$

Each player-season receives the inverse-variance weight, so that larger weights are assigned to more precisely estimated player-seasons.

Specifically, I^2 captures how much of the total variance in performance stems from genuine differences between players, rather than random noise (Higgins & Thompson, 2002). In line with Holzmeister & Johannesson (2025), we treat I^2 as a measure of player skill, with the remaining variance attributed to luck. The I^2 statistic is defined as:

$$I^2 = \tau^2 / (\tau^2 + \sigma^2) \quad (3)$$

where σ^2 denotes the average within-study sampling variance, estimated from the standard errors of the individual player-seasons. If within-season correlation is present, I^2 may be overstated; the reported skill shares should therefore be interpreted as possible upper-bound estimates.

The meta-analytic method aims to identify systematic differences between players. To examine whether a player who overperforms in season *t* also overperforms in season *t*+1, we employ a persistence regression. The persistence regression takes the form:

$$OP_{i,t+1} = \alpha + \beta \times OP_{i,t} + \varepsilon_{i,t+1} \quad (4)$$

where the sample consists of consecutive-season pairs for the same player within the same competition. For players meeting this criterion in two consecutive seasons, we create a pair of observations (OP_t, OP_{t+1}). The coefficient β captures the degree to which overperformance in t predicts overperformance in t+1. If $\beta > 0$, players who outperform expectations in one season systematically do so again; $\beta = 0$ would suggest overperformance is essentially random across time.

If we assume a player's ability is time-stable across seasons, β should converge to I^2. Alternatively, if skill instead follows a first-order autoregressive (AR (1)) process, with autocorrelation parameter ρ, in which a player's skill in t+1 depends on their skill in t, then $\beta = \rho \times I^2$. This connection allows an understanding of the magnitude as well as the persistence of a player's individual finishing and shot-stopping skill.

5. Empirical Results

In total our dataset contains 423,227 outfield player observations and 28,233 goalkeeper observations on match level. After clustering the data in player-seasons and removing player-seasons that do not meet our eligibility criteria, the outfield sample comprises 3,202 player-season observations spanning eight seasons, and the goalkeeper sample includes 932 observations. Table 1 summarizes the descriptive statistics for the outfield sample. The average finishing overperformance is 0.005 and, therefore, almost zero, indicating that the xG model is well calibrated within our sample.

	N	Mean	SD	Min	Max
n_games	3,202	25.867	6.852	10	38
sum_npxG	3,202	6.211	3.430	3	29.600
sum_G	3,202	6.409	4.299	0	33
mean_OP	3,202	0.005	0.094	−0.358	0.465
sd_OP	3,202	0.402	0.125	0.094	0.941

Note: mean_OP is the average of non-penalty goals minus npxG per game within a player-season. Eligibility: ≥10 games with ≥30 min played and cumulative npxG ≥ 3.

Table 1: Descriptive Statistics – Eligible Outfield Player-Seasons

Table 2 reports the descriptive statistics for goalkeepers. The average shot-stopping overperformance is negative (mean_GK_OP = −0.073) and also close to zero, however, the mean deviates more from zero than for outfield players. Therefore, the outfield xG model appears slightly better calibrated than the PSxG model for goalkeepers.

	N	Mean	SD	Min	Max
n_games	932	27.667	8.681	10	38
sum_PSxG	932	36.309	13.577	6.800	77.10
sum_GA	932	38.047	14.658	6	91
mean_GK_OP	932	−0.073	0.183	−0.800	0.570
sd_GK_OP	932	0.781	0.157	0.375	1.560

Note: GK_OP = PSxG − GA per game. Eligibility: ≥10 games with ≥45 min played and cumulative PSxG ≥ 3.

Table 2: Descriptive Statistics – Eligible Goalkeeper-Seasons

We have divided the following empirical results into three distinct parts. The first part is the season-by-season random-effects meta-analysis for the outfield finishing overperformance. The second part is the corresponding analysis for overperformance in goalkeeper shot-stopping. The third part consists of persistence regressions for the two groups.

Table 3 reports the outfield finishing overperformance by season for the results of the REML meta-analysis. Between-player heterogeneity (τ^2) varies by season, with τ ranging from 0.018 (2020/21) to 0.045 (2017/18). The I^2 statistic, which captures the proportion of total variance that is attributed to differences between players, varies from 6.753% (2020/21) to 29.945% (2017/18), with a cross-season average of approximately 21%. Between-player heterogeneity

is genuine and is statistically confirmed by rejecting the null hypothesis of homogeneity in all eight seasons of Cochran's Q test at the 5% level.

Season	k	Pooled OP	τ^2	τ	I^2 (%)	Q	p(Q)
2017/18	410	−0.005	0.002	0.045	29.945	599.130	<0.001
2018/19	416	−0.025	0.001	0.029	15.845	497.169	0.003
2019/20	393	−0.020	0.002	0.041	27.951	533.830	<0.001
2020/21	401	−0.007	<0.001	0.018	6.753	452.285	0.036
2021/22	394	−0.006	0.001	0.035	20.815	489.240	0.001
2022/23	413	−0.013	0.001	0.036	23.156	526.209	<0.001
2023/24	395	−0.017	0.002	0.039	26.824	540.528	<0.001
2024/25	380	−0.005	0.001	0.035	20.638	459.155	0.003

Note: k = number of eligible player-seasons. Pooled OP is the REML-weighted estimate. τ^2 and τ capture between-player heterogeneity (skill). I^2 = share of total variance due to heterogeneity. Q tests homogeneity.

Table 3: Meta-Analysis Results by Season – Outfield Players (REML)

Every season shows between-player heterogeneity that significantly exceeds zero. An average I^2 of about 21% suggests that approximately one-fifth of the total variance in player-level finishing overperformance is due to systematic between-player differences, while the remaining four-fifths is residual variation. The 2020/21 season stands out with the lowest I^2 (6.753%) and τ (0.018). The pooled OP is slightly negative in every season (−0.005 to −0.025). Figure 1 displays the distribution of overperformance in finishing by season.

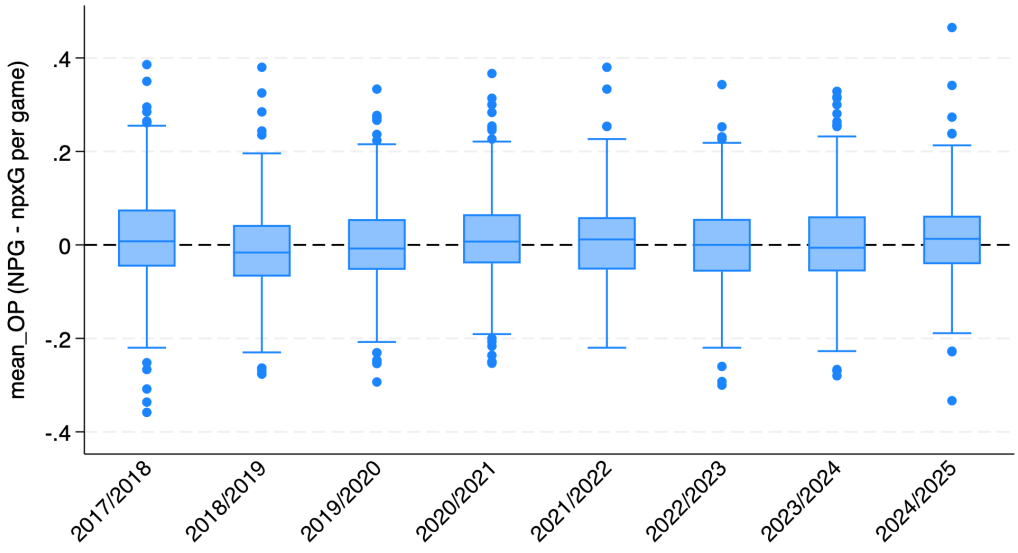

Figure 1: Distribution of Finishing Overperformance (OP) by Season

Player	Sample OP	NPG	npxG	Games	Seasons
Dries Mertens	0.170	53	35.5	123	5
Phil Foden	0.149	53	34.6	120	5
Lionel Messi	0.134	131	102.3	191	6
Luis Muriel	0.119	43	31.2	99	5
Martin Terrier	0.114	51	35	139	6
Florian Sotoca	−0.096	19	32.7	148	5
Edin Džeko	−0.098	69	84.3	164	6
Neal Maupay	−0.099	30	42.1	129	6
Dominic Calvert-Lewin	−0.117	45	64.7	182	8
Christian Kouamé	−0.126	9	18.2	84	4

Note: Sample OP = mean overperformance (NPG − npxG per game) averaged across eligible seasons. Sample restricted to players with ≥4 eligible seasons.

Table 4: Top and Bottom 5 Outfield Players by Sample Overperformance

For a better understanding of the players contributing to the meta-analytic estimates, Table 4 shows the top and bottom five outfield players with the highest sample-average overperformance, considering only players with four or more eligible seasons.

Season	k	Pooled OP	τ^2	τ	I^2 (%)	Q	p(Q)
2017/18	121	−0.038	0.008	0.089	29.633	161.813	0.007
2018/19	123	−0.031	0.010	0.101	34.344	182.389	<0.001
2019/20	114	−0.054	0.009	0.095	31.265	168.445	0.001
2020/21	116	−0.093	0.005	0.068	21.075	138.811	0.065
2021/22	94	−0.066	0.002	0.040	7.690	103.266	0.219
2022/23	127	−0.069	0.011	0.103	32.938	187.837	<0.001
2023/24	117	−0.054	0.008	0.088	26.075	158.613	0.005
2024/25	120	−0.017	0.003	0.057	14.586	139.190	0.100

Note: k = number of eligible goalkeeper-seasons. Pooled GK_OP = REML-weighted estimate. τ^2 = between-GK variance (skill). I^2 = share of total variance due to heterogeneity. Q = Cochran's heterogeneity statistic.

Table 5: Meta-Analysis Results by Season – Goalkeepers (REML)

Table 5 presents the REML meta-analysis results for shot-stopping overperformance. The pooled GK_OP is negative across all seasons, from -0.093 to -0.017, which is consistent with the previously mentioned slightly negative bias in the PSxG model. The between-goalkeeper

heterogeneity (τ) ranges from 0.040 (2021/22) to 0.103 (2022/23), and I^2 ranges from 7.690% (2021/22) to 34.344% (2018/19). The cross-season average I^2 for goalkeepers is approximately 25%, which is about 4 percentage points higher than the average of outfield players. Cochran's Q test rejects homogeneity at or below the 10% level in seven of eight seasons.

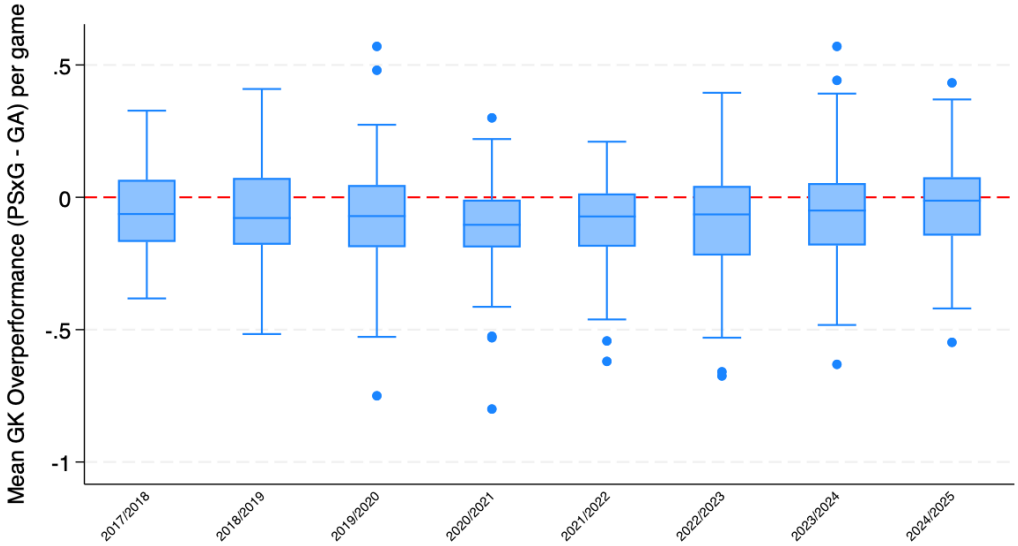

Figure 2: Distribution of Goalkeeper Overperformance (GK_OP) by Season

Across most seasons, goalkeepers show significant heterogeneity in shot-stopping performance. The 2021/22 season stands out with the least heterogeneity shown by the lowest I^2 (7.690%) and τ (0.040). Looking at the two groups, goalkeepers have higher absolute τ values (0.040–0.103) than outfield players (0.018–0.045), indicating that the real between-player dispersion is larger for goalkeepers. The comparable I^2 are therefore explained by a higher within-goalkeeper than within-outfield-player noise. Figure 2 illustrates the distribution of overperformance in shot-stopping by season.

Table 6 lists the five goalkeepers with the highest sample-average shot-stopping overperformance and five with the lowest performance. Only goalkeepers with four or more eligible seasons were included. Brice Samba holds the highest rank with sample GK_OP = 0.214 across 4 seasons. The most consistent underperformer is Lorenzo Montipò (−0.280).

Goalkeeper	Sample GK_OP	GA	PSxG	Games	Seasons
Brice Samba	0.214	158	188.2	140	4
Neto	0.138	159	177.2	125	4
Walter Benítez	0.132	154	169.9	126	4
Mattia Perin	0.086	131	135.8	101	4
Keylor Navas	0.770	107	116	101	5
Alexander Schwolow	−0.220	292	257.3	161	6
Baptiste Reynet	−0.227	238	211.8	129	4
Alex McCarthy	−0.231	189	163.3	118	5
Andrea Consigli	−0.252	411	349.5	248	7
Lorenzo Montipò	−0.280	297	246.1	182	5

Note: Sample GK_OP = mean (PSxG − GA) per game averaged across eligible seasons. Sample restricted to goalkeepers with ≥4 eligible seasons.

Table 6: Top and Bottom 5 Goalkeepers by Sample Overperformance

The outfield players' persistence regressions are summarised in Table 7 with 1,421 matched pairs of consecutive seasons. In the baseline ordinary least squares (OLS) regression with robust standard errors (column 1), the persistence coefficient is $\beta = 0.108$ indicates a positive association between overperformance in season t and overperformance in the subsequent season. Adding season and competition fixed effects and clustering standard errors by player (column 2) yields a coefficient of $\beta = 0.110$ ($p < 0.01$). Finishing overperformance thus shows statistically significant persistence across consecutive seasons.

	(1) OLS	(2) OLS + FE
Persistence (β)	0.108***	0.110***
	(0.027)	(0.030)
Season FE	No	Yes
Competition FE	No	Yes
SE	Robust	Cluster (player)
R^2	0.012	0.021
Observations	1,421	1,421

*Note: Dependent variable: OP in season t+1. OP = mean (NPG − npxG) per game. Sample: matched consecutive-season pairs, both meeting eligibility criteria. FE = fixed effects; SE = standard errors. *** p < 0.01, ** p < 0.05, * p < 0.10.*

Table 7: Persistence of Outfield Player Overperformance

	(1) OLS	(2) OLS + FE
Persistence (β)	0.205***	0.185***
	(0.048)	(0.050)
Season FE	No	Yes
Competition FE	No	Yes
SE	Robust	Cluster (player)
R^2	0.038	0.071
Observations	493	493

*Note: Dependent variable: GK_OP in season t+1. GK_OP = mean (PSxG − GA) per game. Sample: matched consecutive-season pairs, both meeting eligibility criteria. FE = fixed effects; SE = standard errors. *** p < 0.01, ** p < 0.05, * p < 0.10.*

Table 8: Persistence of Goalkeeper Overperformance

Table 8 reports the corresponding persistence regression for goalkeepers. Here, the persistence coefficient is considerably higher: $\beta = 0.205$ ($p < 0.01$) in the baseline specification (column 1) and $\beta = 0.185$ ($p < 0.01$) when fixed effects and clustered standard errors are included (column 2). Both cases report a higher R^2 (0.038 and 0.071), indicating a stronger temporal signal for goalkeeper overperformance than for outfield player finishing. The analysis includes 493 matched pairs of goalkeepers and seasons.

As robustness checks, we vary the eligibility thresholds (outfield: npxG ≥ 5, ≥ 15 games; goalkeepers: PSxG ≥ 5, ≥ 15 games) and conduct league-stratified meta-analyses pooling all seasons within each competition. Under the stricter thresholds, I^2 estimates remain in a similar range to the baseline (outfield: 9.163–37.811%, mean 22.758%; goalkeepers: 5.728–37.085%, mean 25.850%). In the league-stratified analyses, Cochran's Q rejects homogeneity at the 5% level in all five competitions for outfield players, with I^2 ranging from 13.113% (La Liga) to 29.867% (Serie A). For goalkeepers, heterogeneity is significant in four of five leagues, with I^2 ranging from 10.247% (La Liga, p(Q) = 0.370) to 43.491% (Ligue 1).

6. Discussion

The results provide a detailed analysis of individual ability in football performance using our xG-based overperformance metrics. In this section, we discuss our results and their implications for each of our hypotheses, regarding existing literature, limitations, and directions for future research.

H1: Individual outfield players differ systematically in their finishing ability, such that over-performance relative to expected goals is not purely attributable to random variation.

The evidence generally supports H1. The I^2 statistic exceeds zero in all seasons, and Cochran's Q test rejects homogeneity in all eight seasons for outfield players at the 5% level. Thus, some systematic between-player differences explain a modest share of the total variance in player-level finishing overperformance. This finding extends the team-level evidence of Holzmeister & Johannesson (2025) to the player level and is also consistent with Baron et al. (2024), who demonstrate measurable individual differences in shooting accuracy. The 2020/21 season is an outlier with the lowest I^2 (6.753%) which may reflect the crowd-free matches during the COVID-19 period. In the data, the npxG models tend to slightly understate the complexity of scoring which is shown by a pooled OP that is slightly negative every season. One possible explanation is that contextual factors such as defensive pressure and body orientation of the attacker might not be fully captured in the model.

The outfield player-level I^2 (~21%) that we find is lower than the team-level offensive overperformance I^2, which averages approximately 27% across the three seasons reported by Holzmeister & Johannesson (2025). Comparing I^2 estimates across studies indicates a directional tendency but cannot be interpreted as direct equivalence. Our metrics differ in level of aggregation, in their treatment of penalty kicks, and in the underlying xG source. Lower I^2 values at player level are logically plausible. Team-level metrics pool the skills of multiple players into a collective measure, which helps cancel out individual noise and yields a cleaner read on underlying ability. We observe Lionel Messi (sample OP = 0.134 in six seasons) and Harry Kane (0.111 in eight seasons) that are known for their finishing abilities among the best finishers in our sample, which provides a face-validity check on the ranking results.

By definition, the I^2 statistic captures the share of total variance attributable to between-player heterogeneity. Heterogeneity can have other sources besides pure finishing skill, such as persistent tactical roles, positional assignments, or differences in assist quality. Furthermore, our metric captures finishing quality conditional on a shot being taken and does not reflect a player's ability to create or reach shooting positions. This intentional scope re-striction means that our I^2 estimates should be interpreted as a lower bound on the total individual skill involved in goal-scoring. As Davis & Robberechts (2024) show, xG training data can generate persistent biases. It is likely that elite finishers are overrepresented in the model's training set as they tend to shoot from different positions than weaker finishers. This can increase the xG values for similar shot situations, which reduces the overperformance we measure for the best players. In

our case, this bias would deflate the observed between-player differences in overperformance and consequently underestimate the true I^2. However, other model limitations, e.g. ignoring context like defensive pressure or body orientation (Anzer & Bauer 2021), could counter the bias effects in the opposite direction and inflate between-player variation. The net effect of these opposing biases on I^2 is therefore ambiguous. Our reported I^2 estimates should therefore be understood as conditional on the xG model we use, rather than interpreting them as absolute measures of finishing skill.

H2: Individual goalkeepers differ systematically in their shot-stopping ability, such that overperformance relative to post-shot expected goals is not purely attributable to random variation.

We find support for H2, even though the statistical verification is clearer for outfield players. For goalkeepers, Cochran's Q test rejects homogeneity in five of eight seasons at the 5% level and in seven seasons at or below the 10% level. Only the 2021/22 season fails to reach conventional significance. Even though this outlier is not the same season outlier as for outfield players, it still is one of the seasons affected by the COVID-19 pandemic. We observe a systematic negative pooled GK_OP across all seasons, indicating that goalkeepers on average concede slightly more goals than the PSxG model predicts. Our eligibility criteria or the missing inclusion of goalkeeper positioning as mentioned by De-la-Cruz-Torres et al. (2025) could explain this bias. The negative pooled GK_OP does not suggest that the between-goalkeeper I^2 estimates are invalid because the meta-analysis estimates the heterogeneity in the dispersion of player-season effects around the pooled mean, and not the mean itself. A uniformly biased benchmark simply shifts all goalkeeper estimates the same amount and therefore preserves the decomposition of variance. Our average goalkeeper I^2 (~25%) is close to the team-level defensive overperformance I^2 (~24%) reported by Holzmeister & Johannesson (2025). As mentioned before, comparisons between our studies remain directional given the differences in aggregation level, penalty treatment, and data source. The rankings of goalkeepers present another way to conduct a face validity check. Brice Samba (sample GK_OP = 0.214 over four seasons) and Keylor Navas (0.077) are among the best shot-stoppers, while Andrea Consigli (−0.252) and Alexander Schwolow (−0.220) are among the worst.

H3: Goalkeepers exhibit a more pronounced skill component in shot-stopping than outfield players exhibit in finishing.

The evidence for H3 is not highly pronounced and depends on the metric we consider. The average I^2 for goalkeepers (~25%) is slightly higher than for outfield players (~21%), but both

lie in a similar range, suggesting that the share of between-player variance is quite similar between the two groups. However, regarding absolute values, the between-player standard deviation τ is systematically higher for goalkeepers (0.040–0.103) than for outfield players (0.018–0.045). This provides an indication that the real skill differences in shot-stopping among the goalkeepers group is substantially larger than the differences in finishing for outfield players. A correspondingly higher within-goalkeeper noise explains the relatively similar I^2 values. A structural explanation for this difference is that goalkeepers face shots as externally generated events, whereas outfield players must first create or reach a shooting position – a skill dimension that is not captured by xG-based overperformance and thus adds unobserved heterogeneity to the outfield sample. Goalkeeper differences are also more stable over time, which provides an additional signal for skill and will be discussed in the following.

H4: Individual finishing and shot-stopping overperformance persists across seasons, indicating the presence of a stable, skill-based component in player performance.

We can confirm H4 due to our persistence regression results. Both the outfield and goalkeeper samples show that performance persists across seasons as the persistence coefficients are positive and highly significant. This result is consistent with the persistence logic formalised by Miles et al. (2013) and Getty et al. (2018): if overperformance was random, there would be no serial correlation. The persistence coefficient for goalkeepers ($\beta = 0.185$) is notably larger than for outfield finishers ($\beta = 0.110$), and the explained variance is also higher ($R^2 = 0.071$ vs. 0.021) in both of our specifications (with or without fixed effects), indicating that goalkeeper overperformance is more predictable across seasons. The positive and significant coefficients complement our meta-analytic I^2 results. As discussed at the end of Section 4, under time-constant skill β would converge to I^2. The observed β values (0.110 for outfield, 0.185 for goalkeepers) between zero and our average I^2 estimates (~21% and ~25%). We therefore assume that skill is not perfectly time-constant but follows an autoregressive process. As elaborated in Section 4, in a first-order autoregressive process with autocorrelation parameter ρ, the persistence coefficient takes the form $\beta = \rho \times I^2$. We therefore calculate ρ values of approximately 0.52 for outfield players and 0.74 for goalkeepers implied under the AR(1) assumption and using average I^2 values. These implied autocorrelation parameters have a direct interpretation: for outfield players finishing skill retains only about half its signal from one season to the next, while a goalkeeper's shot-stopping skill retains roughly three-quarters. The season-based systematic component of individual performance that we identified with our meta-analysis therefore decays substantially over time. This implies that single season observations are highly

unreliable indicators to separate overperformance from underlying ability. Observing a higher ρ for goalkeepers provides additional support for H3 because shot-stopping as a skill is not only more dispersed but more stable over time. The directional consistency between the methods (meta-analysis and persistence regression) provides evidence of internal coherence.

Our persistence estimates align with independent evidence as Kharrat et al. (2020) report season-to-season correlations of 0.35 using plus-minus player ratings, which, while not directly comparable to our results due to methodological differences, suggests that individual performance contains a stable but modest skill signal that carries over across seasons. Persistence results should be contextualised with explained variance. $R^2 = 0.012$ means current-season overperformance for outfield players explains roughly 1.2% of variance in next-season overperformance. For goalkeepers, $R^2 = 0.038$ implies roughly 3.8% variance. These values do not negate the finding of skill because the coefficients, albeit low, are statistically significant and economically relevant. However, the coefficients show most of between-season variation in individual performance is unpredictable, even with the presence of skill.

We derive practical implications of these results. Regarding scouting and player valuation, results indicate caution when interpreting short run overperformance: an outlier individual season is more likely the result of luck than a stable signal of skill, and identifying true talent requires multiple season observations. Because most observed overperformance is luck-based, a club or analyst looking purely at one season xG overperformance will be choosing mostly on noise. This has direct implications for data-driven player valuation. Müller, Simons & Weinmann (2017) show that statistical models that include individual performance data, such as goals and assists, can value outfield players as accurately as valuations based on crowd-sourced data from Transfermarkt. However, if a large share of the observed goal-scoring output is the result of luck rather than skill, single season performance measures can bias market-value estimations by overvaluing players whose valuations were influenced by positive luck and undervaluing players whose valuations were not influenced by positive luck. This concern is directly corroborated by Gauriot & Page (2019), who use shots hitting the post as a quasi-experimental setting to illustrate that players' evaluations are influenced by random outcomes causally: managers give players who score in as-good-as-random situations far more playing time and expert journalists assign them substantially higher ratings in violation of the informativeness principle. The decision-makers in their study overweight observed outcomes with respect to their informational value about actual talent, which is the distortion our decomposition aims to quantify. This paper offers a calibrated basis for adjusting performance-based evaluations for the noise

component by providing season-by-season estimates of the skill share based on individual-level xG overperformance data. Given that we attribute 66–93% of individual overperformance variance to luck (I^2 estimates of 6.753–34.344% across seasons), wage premia and player values based on single season output are very susceptible to mispricing.

From a personnel economics perspective, the informativeness principle (Holmström 1979) states that optimal contracts should condition compensation only on signals that carry information about agent effort or ability. Our results suggest that evaluating players based on their single season xG overperformance is based on limited information about underlying player ability due to the high amount of variance attributable to luck. Performance-contingent contracts based on such output measures therefore systematically misallocate resources, rewarding randomness rather than skill (Prendergast 1999). Even though we cannot test these contractual mechanisms directly due to a lack of contract data, we provide the empirical signal-to-noise ratios that calibrate the theoretical prediction.

Several limitations should be acknowledged when interpreting the results of this research. First, we collect our data from one specific xG model (Opta via FBref) whose exact specification is not publicly documented. If the model is biased in calibration for certain shot types, player types, or tactical contexts, this could strongly influence our overperformance measures. Davis & Robberechts (2024) show training-data interdependencies of elite finishers by as much as 17% in the case of Lionel Messi, which suggests a likelihood that the overperformance we measure underestimates elite finishers. This constraint is not specific to our study. Holzmeister & Johannesson (2025), whose methodology we extend, similarly rely on a commercial xG model (Football xG). The dependence on proprietary xG models is thus an inherent limitation of the empirical literature in this field. Second, by defining eligibility criteria we inevitably introduce selection effects, as we exclude players with limited playing time in attacking roles or in general. The cumulative npxG threshold disproportionately excludes defenders, so that our sample of outfield players is dominated by attacking players. Third, when forming the persistence sample we introduce a survivorship bias: we only include players into the regression who meet the eligibility criteria in two consecutive seasons within the same competition. This reduces the sample size but improves internal comparability, since overperformance values are more directly comparable within a single competitive context. It excludes high-performing players who transfer to stronger leagues and underperforming players who move to weaker leagues. Both influence β with opposite directional effect and have therefore countervailing implications for the estimated persistence coefficient. Fourth, the analysis does not control for

age or career phase. Younger players may be improving within our observation period, and older players may be declining, which could affect the persistence coefficient. Incorporating age profiles would be a valuable extension but exceeds the scope of our analysis. Fifth, the precision weights are derived from game-count-based standard errors instead of shot-level measures. Consequently, players with similar game counts but substantially different shot volumes could be imprecisely weighted.

Future research could address the limitations identified above in several ways. First, recreating our analysis with data from a different xG data provider would help assess the sensitivity of the decomposition to model specification. Second, extending the analysis to additional performance metrics, such as expected assists, could provide a more comprehensive picture of individual skill. However, this extension would raise an identification problem. An assist is only recorded if the receiver scores, so the observed outcome depends partly on the finisher and therefore contains additional luck. Third, the persistence analysis could be extended by introducing a within-season split, following Getty et al. (2018). Fourth, the baseline analysis does not estimate league-specific models; league-stratified analyses are included only as robustness checks. Future work could further examine whether there are systematic differences across leagues which would reveal whether the competitive environment affects the detectability of individual finishing and shot-stopping ability.

7. Conclusion

This paper extends the team-level decomposition of skill and luck from Holzmeister & Johannesson (2025) to the individual player level. To our knowledge, we provide the first systematic meta-analytic estimate of between-player heterogeneity in finishing and shot-stopping overperformance in European football by using REML-based random-effects meta-analysis. We also extend the approach by adding a complementary persistence analysis across eight seasons of data from Europe's five major leagues. Our analysis makes three main contributions. First, we find that individual-level skill in both finishing and shot-stopping is statistically detectable but modest: I^2 estimates of 6.753–34.344% across seasons imply that most of the season-level performance heterogeneity between players can be attributed to luck rather than ability. Second, the player-level heterogeneity shares are somewhat lower than the team-level shares reported by Holzmeister & Johannesson (2025). Apart from differences in collected data, this is a structural consequence of the lower aggregation inherent in individual-level measurement, rather than a substantive indication of less skill. Third, the persistence analysis shows that over-

performance stays partly stable across consecutive seasons ($\beta = 0.110$ for outfield, 0.185 for goalkeepers). Because $\beta = \rho \times I^2$, dividing the fixed-effects persistence estimates by the cross-season average I^2 (21% for outfield, 25% for goalkeepers) allows us to recover implied season-to-season correlations of underlying skill under the AR(1) assumption: approximately 52% of the skill signal is stable across consecutive seasons for outfield players, and 74% for goalkeepers. Single season xG overperformance, our estimates suggest, is too noisy to serve as a reliable signal of player ability; robust evaluation calls for multi-season evidence, shrinkage toward the mean, and explicit accounting for the noise components quantified here.

Literature

Anzer, Gabriel/Bauer, Pascal (2021): "A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer)", Frontiers in Sports and Active Living 3, 624475.

Baron, Ethan/Sandholtz, Nathan/Pleuler, Devin/Chan, Timothy C. Y. (2024): "Miss it like Messi: Extracting Value from Off-Target Shots in Soccer", Journal of Quantitative Analysis in Sports 20(1), pp. 37–50.

Borenstein, Michael/Hedges, Larry V./Higgins, Julian P. T./Rothstein, Hannah R. (2010): "A Basic Introduction to Fixed-Effect and Random-Effects Models for Meta-Analysis", Research Synthesis Methods 1(2), pp. 97–111.

Brechot, Marc/Flepp, Raphael (2020): "Dealing with Randomness in Match Outcomes: How to Rethink Performance Evaluation in European Club Football Using Expected Goals", Journal of Sports Economics 21(4), pp. 335–362.

Davis, Jesse/Robberechts, Pieter (2024): "Biases in Expected Goals Models Confound Finishing Ability", arXiv preprint arXiv:2401.09940.

De-la-Cruz-Torres, Blanca/Navarro-Castro, Miguel/Ruiz-de-Alarcón-Quintero, Anselmo (2025): "An Expected Goals on Target (xGOT) Model: Accounting for Goalkeeper Performance in Football", Big Data and Cognitive Computing 9(3), 64.

Gauriot, Romain/Page, Lionel (2019): "Fooled by Performance Randomness: Overrewarding Luck", Review of Economics and Statistics 101(4), pp. 658–666.

Getty, Daniel/Li, Hao/Yano, Masayuki/Gao, Charles/Hosoi, A. E. (2018): "Luck and the Law: Quantifying Chance in Fantasy Sports and Other Contests", SIAM Review 60(4), pp. 869–887.

Heuer, Andreas/Müller, Christian/Rubner, Oliver (2010): "Soccer: Is Scoring Goals a Predictable Poissonian Process?", Europhysics Letters (EPL) 89(3), 38007.

Higgins, Julian P. T./Thompson, Simon G. (2002): "Quantifying Heterogeneity in a Meta-Analysis", Statistics in Medicine 21(11), pp. 1539–1558.

Holmström, Bengt (1979): "Moral Hazard and Observability", Bell Journal of Economics 10(1), pp. 74–91.

Holzmeister, Felix/Johannesson, Magnus (2025): “Skills vs. Luck: Decomposing Deviations from Expected Performance in European Football Leagues”, Journal of Sports Economics 26(8), pp. 953–975.

Kharrat, Tarak/McHale, Ian G./López Peña, Javier (2020): “Plus–Minus Player Ratings for Soccer”, European Journal of Operational Research 283(2), pp. 726–736.

Lames, Martin (2018): “Chance Involvement in Goal Scoring in Football: An Empirical Approach”, German Journal of Exercise and Sport Research 48(2), pp. 278–286.

Miles, Thomas J./Levitt, Steven D./Rosenfield, Andrew M. (2013): “Is Texas Hold ‘Em a Game of Chance? A Legal and Economic Analysis”, Georgetown Law Journal 101, pp. 581–636.

Mauboussin, Michael J. (2012): “The Success Equation: Untangling Skill and Luck in Business, Sports, and Investing”, Harvard Business Review Press, Boston, MA.

Müller, Oliver/Simons, Alexander/Weinmann, Markus (2017): “Beyond Crowd Judgments: Data-Driven Estimation of Market Value in Association Football”, European Journal of Operational Research 263(2), pp. 611–624.

Prendergast, Canice (1999): “The Provision of Incentives in Firms”, Journal of Economic Literature 37(1), pp. 7–63.

Rathke, Alex (2017): “An Examination of Expected Goals and Shot Efficiency in Soccer”, Journal of Human Sport and Exercise 12(2proc), pp. S514–S529.

Robberechts, Pieter/Davis, Jesse (2020): “How Data Availability Affects the Ability to Learn Good xG Models”, in: Brefeld, Ulf/Davis, Jesse/Van Haaren, Jan/Zimmermann, Albrecht (eds.), Machine Learning and Data Mining for Sports Analytics (MLSA 2020), Communications in Computer and Information Science (CCIS), vol. 1324, Springer, Cham, pp. 17–27.

Sarkar, Sumit/Kamath, Sooraj (2023): “Does Luck Play a Role in the Determination of the Rank Positions in Football Leagues? A Study of Europe’s ‘Big Five’”, Annals of Operations Research 325(1), pp. 245–260.

Viechtbauer, Wolfgang (2005): “Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model”, Journal of Educational and Behavioral Statistics 30(3), pp. 261–293.

Wunderlich, Fabian/Seck, Alessandro/Memmert, Daniel (2021): “The Influence of Randomness on Goals in Football Decreases over Time: An Empirical Analysis of Randomness Involved in Goal Scoring in the English Premier League”, Journal of Sports Sciences 39(20), pp. 2322–2337.

Diskussionspapiere des Instituts für Organisationsökonomik

Seit Institutsgründung im Oktober 2010 ist monatlich ein Diskussionspapier erschienen. Im Folgenden werden die letzten zwölf aufgeführt. Eine vollständige Liste mit Downloadmöglichkeit findet sich unter http://www.wiwi.uni-muenster.de/io/forschen/diskussionspapiere.html

DP-IO 3/2026 Individual Skill or Just Luck?
Decomposing Individual Finishing and Shot-Stopping Performance in European Football
Tom Böttger/Lars Vischer
März 2026

DP-IO 2/2026 Leitfaden für Seminare am Institut für Organisationsökonomik
Alexander Dilger
Februar 2026

DP-IO 1/2026 Home Bias and Referee Bias in European Football
The Role of Stoppage Time Allocation
Tom Böttger/Lars Vischer
Januar 2026

DP-IO 12/2025 Formen und Risiken politischer Brandmauern
Alexander Dilger
Dezember 2025

DP-IO 11/2025 Flexibler Übergang in den Ruhestand
Alexander Dilger
November 2025

DP-IO 10/2025 15. Jahresbericht des Instituts für Organisationsökonomik
Alexander Dilger
Oktober 2025

DP-IO 9/2025 Abschlussarbeiten am Institut für Organisationsökonomik
Alexander Dilger
September 2025

DP-IO 8/2025 Gesamtliste zum VHB Rating 2024
Alexander Dilger
August 2025

DP-IO 7/2025 Betreuung und Begutachtung von Dissertationen vereint oder getrennt
Alexander Dilger
Juli 2025

DP-IO 6/2025 Erfahrungen und Planungen zu KI in Lehre und Leistungsbeurteilungen
Alexander Dilger
Juni 2025

DP-IO 5/2025 Home Bias in Different German Football Leagues
Tom Böttger/Lars Vischer
Mai 2025

DP-IO 4/2025 Prospect Theory and Psychological Factors in Football
Effects of the Video Assistant Referee on Penalty Conversion Rates in the German Bundesliga
Tom Böttger/Lars Vischer
April 2025

Herausgeber:
Prof. Dr. Alexander Dilger
Universität Münster
Institut für Organisationsökonomik
Scharnhorststr. 100
D-48151 Münster

Tel: +49-251/83-24303
Fax: +49-251/83-28429

www.wiwi.uni-muenster.de/io