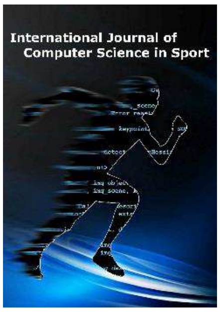
International Journal of Computer Science in Sport

sciendo

Volume 22, Issue 2, 2023

Journal homepage: http://iacss.org/index.php?id=30

DOI: 10.2478/ijcss-2023-0013



Success-Score in Professional Soccer – Is there a sweet spot in the analysis of space and ball control?

David Brinkjans, Daniel Memmert, Yannik Paul, & Jürgen Perl

German Sports University Cologne

Abstract

In contrast to simple performance indicators in the practical application of quantitative analysis in professional soccer, the inclusion of certain contextual elements can improve both the predictive quality and interpretability of these. Therefore, the Success-Score is intended to identify the factors relevant to success by linking ball control and space control.

Position datasets from 14 games of the Bundesliga were used to calculate Success-Scores for several interval lengths for the penalty area and the 30-meter-zone. The relative goalscoring frequency above resp. below the 80th percentile, the rank correlation in terms of goals scored pursuant to the sorting of the Success-Score as well as possible distinctions in the Success-Score between two teams of different quality were examined.

Results revealed that interval lengths and the area under investigation largely affect the resulting Success-Score and its distribution. The Success-Score applied to the 30-meter-zone seems preferable when analyzing goalscoring. Dependent on the target of analysis, methodological and theoretical considerations need to be balanced in a sweet spot of the interval length.

KEYWORDS: KEY PERFORMANCE INDICATOR, BALL POSSESSION, DOMINANT REGIONS, FOOTBALL, EFFICIENCY.

Introduction

Professional sport is one of the most competitive areas in our society. Soccer is no exception (Memmert, 2021). Every team wants to achieve the maximum. Hence, every little edge that can be gained is relevant. One of many directions in which this has led is Key Performance Indicators (KPIs) and quantitative match analysis more generally (Low et al., 2017; Memmert & Raabe, 2018; Memmert et al., 2017). With KPIs researchers and practitioners attempt to pinpoint relevant factors that contribute to success (Schlenger et al., 2023). To this end complex behaviors or interactions are reduced to single values as in scales, ratings or rankings (Perl & Memmert, 2017). This approach is not unique to soccer and has been applied in other sports as well (see also Basketball and Cricket: Dehesa et al., 2019; Petersen, 2017). Regardless of the success that several studies have had in linking match actions to success (Sarmento et al., 2014) the approach of using isolated KPIs has its drawbacks. This is the case not least in such complex and interactive sports as soccer (Perl & Memmert, 2018). To a large part, it comes down to the challenge of keeping relevant information and context while cutting down the information to the essential core (Mackenzie & Cushion, 2013; Rein & Memmert, 2016; Sarmento et al., 2014). As a consequence, KPIs based on more extensive contextual information have been developed (Lord et al., 2020). Nevertheless, KPIs require careful interpretation with consideration of the context and their limitations (Memmert & Rein, 2018). Including certain contextual elements in the analysis has been shown to enhance the predictive quality as well as interpretability of KPIs (Phatak et al., 2022). Despite these facts, a recent review showed that the majority of published articles on performance analysis in invasion sports including soccer focus on comparably plain KPIs (Lord et al., 2020). In addition, this trend was also found in a survey with practitioners who reported relying on event data and relatively simple KPIs more frequently than positional data and more complex KPIs (Herold et al., 2021; Memmert et al., 2017; Low et al. 2019).

Ball possession is an element of the game that gets a lot of attention in performance analysis research and has done so for many years (Jones et al., 2004). It seems hardly possible to score goals without possession of the ball and hence it comes as no surprise that a lot of research has focused on the connection between ball possession and success. The evidence is mixed nonetheless. While higher-ranked teams are predominantly reported to have more time in possession of the ball than lower-ranked teams (Collet, 2013; H. Liu, Yi, et al., 2015; Vogelbein et al., 2014), results for match status and match outcome are mixed. Some researchers reported higher possession percentages in winning performances (Lago-Peñas et al., 2011; H. Liu, Gomez, et al., 2015), while others found no differences (Alves et al., 2019; H. Liu, Yi, et al., 2015) or higher possession percentages in losing performances (Collet, 2013). As proposed before (Brinkjans et al., 2022; H. Liu et al., 2016), a possible explanation for these putative inconsistencies is the lack of distinction between within-team and between-team effects. Regardless of team quality or success, when losing all teams are somewhat forced to become more proactive, resulting in higher ball possession percentages (H. Liu et al., 2016; Mao et al., 2016). Nevertheless, as a general trend, better teams tend to have more possession, most likely at least partially as the consequence of a strategic choice (Gollan et al., 2018; Kempe et al., 2014; Zhou et al., 2021). Furthermore, the complex nature of soccer is not to be forgotten in this context and is the reason why one simple parameter will always have limited predictive value regarding success (Memmert & Rein, 2018).

Another common approach is to distinguish between different areas on the pitch. Tenga et al. (2010) found an extended penalty area of 30 meters to be an area in which possession is linked with the number of goals scored. Other researchers were able to link entries into the penalty area with winning performances (T. Liu et al., 2021; Ruiz-Ruiz et al., 2013). More generally, the offensive midfield area has been shown to be the most important area on the pitch for goalscoring (Caicedo-Parada et al., 2020). These results underpin the logical assumption that possession in certain areas on the pitch is more valuable than in other areas in regards to goalscoring but do not clarify how this value might vary within the final third (e.g. penalty area).

Despite these findings, nothing is gained by possession alone. Possession has to be utilized efficiently (Winter & Pfeiffer, 2016). One part of utilizing possession efficiently, though not dependent on possession in itself, is the concept of space control. These so-called dominant regions are estimated by a model and represent the area that a certain player can get to before any other player (Fujimura & Sugihara, 2005; Taki & Hasegawa, 2000). Space control has proven to be a useful concept (Rein et al., 2017; Rein & Memmert, 2016) and has been linked with goalscoring as well as the probability of winning (Rein et al., 2017). While space control is a valuable concept, it essentially lacks one key piece of information from an attacking perspective – ball possession. No matter how much space is controlled by an attacking team, they need to be in possession of the ball and subsequently need to get the ball into the right areas.

With these facts in mind, a dynamic KPI based on the software tool SOCCER© by Jürgen Perl (for further information see: Perl et al., 2013; Perl & Memmert 2011) named Success-Score was developed by Perl & Memmert (2017). The Success-Score aims to bridge the gap left by simple KPIs by combining space control and ball possession in specific areas. The idea is that the combination of the two contributes to or even preconditions goalscoring (Brinkjans et al., 2022; Perl & Memmert, 2017; Memmert et al., 2021). The approach combines space control and ball possession over a certain interval for a predefined area. A previous study has validated the Success-Score for a 300-second interval and the area within 30 meters of the opponent's goal by establishing a link between the Success-Score based on said parameters and goalscoring as well as inter-team differences (Brinkjans et al., 2022). In the development of the Success-Score the authors noted that the optimal interval is dependent on the focus of the analysis but most likely lies somewhere between 100 and 600 seconds (Perl & Memmert, 2017). Logically, the interval length should affect the resulting Success-Score profiles. Longer intervals result in a smoothing of extreme scores since the resulting Success-Scores is an average for a larger number of scores. In contrast, shorter intervals should produce spikier Success-Score profiles. Besides these statistical considerations, the argument is that longer intervals would remove the dynamic of the KPI, while shorter intervals merely replicate the events of the match (Perl & Memmert, 2017). The chosen interval length inevitably interacts with the area of choice. The area of 30 meters tested in the previous investigation has proven a reasonable approach (Brinkjans et al., 2022). Nevertheless, one might argue that 16 meters (i.e. the penalty area) should be in closer relation to success and goalscoring as it is closer to the ultimate target – the goal. For the same reasons, the penalty area is usually a more severely defended area. Hence, instances of ball and space control are expected to be less frequent in the penalty box than within a 30-meter area for the attacking team. This inevitably affects the Success-Score distribution. Therefore, a comparison is warranted. Subsequently, this paper aims to compare different interval lengths as well as a 16-meter-zone (i.e. penalty area) in addition to the previously used 30-meter-zone. It will be addressed whether (1) a sweet spot exists for the parameters resulting in an optimal version of the Success-Score and (2) whether this potential sweet spot varies depending on the aim of the analysis.

Methods

Success-Score

The Success-Score combines space control and ball control. The idea was not to simply add up both parameters. Instead, additional expressiveness is created by constructing Success as the product of Efficiency and Effort. The model's definitions of Effort and Efficiency are based on the well-accepted definitions of efficiency as the ratio of useful output (what we define as "Success") and input (what we define as "Effort") which can then be rewritten as follows:

$$\text{Efficiency} = \text{Success} / \text{Effort} \quad \textit{is equivalent to} \quad \text{Success} = \text{Efficiency} * \text{Effort} \tag{1}$$

On the one hand, Effort was defined as the sum of occurrences of ball possession events (B) and/or space control rates above 20% (S) within a certain space in front of the opponent's goal. This represents the attempts made to achieve ball possession and space control within said area. On the other hand, Efficiency is defined as the correlation between the space control rate (SR) and ball control events (B). This is because both parameters on their own have limited value in the aim of creating goalscoring opportunities. But if they occur in combination, they do. Therefore, the stronger they correlate, the higher the efficiency. This led to the following equation for the Success-Score:

$$Success(t_{start}, t_{end}) = r(SR, B)_{t_{start}}^{t_{end}} * \left(\sum_{t=t_{start}}^{t_{end}} \left(S(t) \oplus B(t) \right) \right) / IL \tag{2}$$

For a detailed description of equation 2 check Perl & Memmert (2017). Based on this equation, the Success-Score should be highest if space control and ball possession values are high and highly correlated and lowest if ball possession and space control are rare and barely or even negatively correlated. While statistically possible, a systematic negative relationship between ball and space control and therefore a negative Efficiency are deemed practically implausible. On the grounds that a negative correlation between ball control and space control is practically no worse than no correlation at all, all negative Success-Scores are set equal to zero.

The Success-Score is calculated for each second (t_{end}) and represents the team's performance in a given time interval prior to that second. If the time point t = 500 seconds (i.e. $t_{end} = 500$) is analyzed, and $\Delta t = t_{end} - t_{start} = 300$ seconds, the Success-Score refers to the interval between the 200th and the 500th second. By calculating one Success-Score for each of two opposing teams and moving t_{end} one second at a time, a dynamic KPI in form of a Success-Score profile is achieved. This dynamic KPI portrays the dynamic interaction of both teams' attacking performance.

Data

The data analysis (link to all relevant scripts in the appendix) is an extension of the validation conducted with 300 seconds and 30 meters and uses the same data (Brinkjans et al., 2022). That data comprised 56 position datasets. 14 Bundesliga matches and hence 28 halftimes with one position dataset for each of the two teams were processed via SOCCER© (Perl et al., 2013; Perl & Memmert, 2011). These position data sets contained x- and y-coordinates for each player involved with a framerate of 1 Hz. The focus lay on two specific teams. Team A – a team with 43% average ball possession and a final table position in the lower midrange of the 18 teams – represented a team with comparably little success and little ball possession. Team B – a team with 55% average ball possession that finished in the top 4 – represented a top team with a possession-heavy playing style. With the help of SOCCER© (Perl et al., 2013; Perl & Memmert, 2011), Success-Scores were calculated at a frequency of 1Hz for a number of interval lengths for 16 meters and 30 meters. In a first step, an initial examination of several interval lengths (100, 150, 200, 300, 400 and 500 seconds) was conducted based on the recommended interval length of 100-600 seconds by Perl & Memmert (2017). The examinations left open questions regarding the effect of the interval length on the Success-Score outside the selected bounds. As a consequence, additional interval lengths were included (10, 25, 50, 750 1000, 1250 and 1500 seconds). Irregularities in the distribution of the Success-Scores for 16 meters and 100 and 150 seconds as well as 30 meters and 200 and 300 seconds warranted the additional inclusion of the interval lengths 110, 120, 130 and 140 seconds for 16 meters and 225, 250 and 275 seconds for 30 meters. The full analysis was therefore conducted on 16 datasets for 16 meters and 15 datasets for 30 meters. At this point, it needs to be clarified that the added interval lengths were not added expecting larger effects in regards to team strength and goalscoring and hence a practical value. Rather they were chosen to aid the interpretation of the interaction of the Success-Score and the interval length upon which it is based.

For each goal scored in the included matches, the exact timecode was determined based on video data with the help of experts (ICC > .80). Thereby each goal was linked to one specific data point. The timecode was relocated back to the second that entailed the origin of the set-piece for goals scored from free-kicks and penalties. The intention was to avoid distortion of the Success-Score by the play-unrelated movement of players between the time of the foul and the execution of the set-piece. Subsequently, with the same video data, goals were categorized based on whether the goal was scored from within or outside the penalty area. The two experts agreed to 100% resulting in a Cohen's kappa of $\varkappa = 1$.

With the intention of analyzing the link between the Success-Score and goalscoring, each dataset was split into two groups based on the Success Score. All data points above resp. below the 80th percentile were scaled nominally (goal/no goal) and the relative goalscoring frequency was compared using the Chi-squared test of independence. For the Success-Scores based on 16 meters only the goals scored from within that area were categorized as goals because there is no reason to assume a connection between goals scored from outside the penalty area and space and ball control within the penalty area. The Chi-squared test is an appropriate choice when comparing the proportions of categorical outcomes (Kirkwood et al., 2003). A priori, the significance level was set at 0.05 which is equivalent to a critical Chi-square value of 3.841 for the analysis of this data. In another step, each dataset was split into equally sized eighths after sorting the datasets by the Success-Score. The number of goals for each eighth was determined. Again, for the 16-meter-Success-Scores only those goals scored from within the box were categorized as goals. Those goals linked to a Success-Score that occurred in more than one of the eighths were divided based on the frequency of the relevant Success-Score in each of the eighths and distributed accordingly. Kendall's Tau-b rank correlation coefficient was utilized to determine whether the Success-Score is linked to goalscoring based on these eighths. Kendall's tau was deemed superior to Spearman's rank correlation coefficient due to the small sample size and frequency of tied ranks (Field et al., 2012; Prematunga, 2012).

Inter-team differences between Team A and Team B were tested with the Wilcoxon-Mann-Whitney-U test due to the deviations from normality and the heterogeneity of variance for most of the datasets which were mainly caused by the large share of zero. The size of the effects was assessed by dividing U by the product of the size of both teams' datasets (i.e. the sample sizes) representing the likelihood of a random data point from one team being higher than a random data point from the other team (Conroy, 2012). Due to the large number of individual data points, the test is largely overpowered. Hence, the calculated effect size will be the primary focus of interpretation.

Results

The distributions of the Success-Score differ largely between the datasets based on the chosen parameters (Table 1). The range, the standard deviation and the proportion of the dataset being zero decrease with an increase in interval length for both 16 and 30 meters. The mean and median increase with an increase in interval length at first but decrease for interval lengths above 300 seconds for both 16 and 30 meters. Across interval lengths, the proportion of zeros is smaller for Success-Scores based on 30 meters. Furthermore, the mean, median, range and standard deviation are larger for 30-meter-Success-Scores.

Table 1: Descriptive statistics for all datasets.

	16 meters					30 meters				
Interval length (s)	Range	% 0	$\bar{x}$	$\tilde{x}$	SD	Range	% 0	$\bar{x}$	$\tilde{x}$	SD
10	0 - 99	91.25%	2.28	0	9.12	0 - 99.8	79.91%	7.42	0	18.60
25	0 - 94.2	81.88%	2.68	0	7.48	0 - 98.6	63.21%	9.67	0	17.59
50	0 - 94.5	69.24%	2.94	0	6.24	0 - 97.9	44.26%	11.07	2.6	15.98
100	0 - 68.8	48.55%	3.38	0.2	5.66	0 - 87	23.19%	12.16	8.1	13.67
110	0 - 57	47.11%	3.03	0.4	4.72					
120	0 - 52.1	44.58%	3.02	0.7	4.55					
130	0 - 47.8	42.06%	3.01	0.9	4.39					
140	0 - 44.9	39.95%	2.99	1.1	4.25					
150	0 - 42.5	37.51%	2.99	1.3	4.12	0 - 84.5	14.67%	12.28	9.4	12.13
200	0 - 36.2	28.36%	2.96	1.7	3.64	0 - 69.2	10.09%	12.39	9.9	11.11
225						0 - 65.2	8.38%	12.40	10.1	10.65
250						0 - 63.9	7.48%	12.37	10.3	10.27
275						0 - 61.4	6.79%	12.34	10.4	9.93
300	0 - 25.6	18.95%	2.86	2	3.05	0 - 58.1	6.21%	12.31	10.5	9.64
400	0 - 20.2	14.22%	2.80	2.2	2.73	0 - 55.2	4.86%	12.18	10.6	8.85
500	0 - 16.6	11.85%	2.73	2.2	2.52	0 - 51.2	4.48%	11.93	10.6	8.27
750	0 - 13.9	9.28%	2.57	2.1	2.18	0 - 47.8	4.17%	11.37	10.5	7.50
1000	0 - 11.7	8.69%	2.40	2.1	1.97	0 - 43.2	4.10%	10.84	10.1	7.14
1250	0 - 9.6	8.62%	2.25	2	1.83	0 - 41	4.12%	10.21	9.6	6.88
1500	0 - 9.4	8.50%	2.10	1.8	1.73	0 - 39.4	4.13%	9.59	9.1	6.65

Note. %0: the percentage share of zero; $\bar{x}$: mean; $\tilde{x}$: median; SD: standard deviation; all values are rounded to two decimal places.

The results of the correlation analysis can be found in Table 2. Due to the large number of tests, the family-wise error rate should be kept in mind. Therefore, the overall trend of the effect sizes should receive similar attention as the p-values. In general, a slight trend of more goals being scored at higher Success-Scores seems apparent in Tables 3 and 4. The rank correlation between the number of goals and the Success-Score was statistically significant for six of the 17 datasets for 16 meters. Although some medium-sized statistically non-significant effects were found for other datasets as well. The correlation coefficient shows a large variation across all interval lengths. Regardless, the correlation coefficients tend to be larger for small to medium interval lengths (up to 500 seconds). The datasets based on 30 meters show a different picture. From 10 to 500 seconds in all but two datasets a statistically significant rank correlation was found. The two exceptions were just barely not significant at the 0.05 level. No statistically significant correlations were detected for the longer intervals. Large effects are found for all correlations up to the 500-second interval.

Table 2: Results of the rank order correlation and the chi-square test.

	16 meters					30 meters				
Interval length (s)	τ	p-value	χ2	U80	O80	τ	p-value	χ2	U80	O80
10	0.5	0.06	5.35*	25.42	13.58	0.7*	0.01	61.27*	15	30
25	0.68*	0.02	2.37	27.36	11.64	0.71*	0.01	13.89*	26	19
50	0.24	0.21	2.83	27	12	0.69*	0.01	5.00*	30	15
100	0.33	0.13	8.31*	24	15	0.69*	0.01	6.81*	29	16
110	0.62*	0.02	8.31*	24	15					
120	0.79*	0.00	6.14*	25	14					
130	0.76*	0.00	3.96*	26.23	12.77					
140	0.79*	0.00	1.64	28	11					
150	0.42	0.08	0.23	30	9	0.5	0.05	1.25	33	12
200	0.4	0.09	4.33*	26	13	0.69*	0.01	5.00*	30	15
225						0.84*	0.00	3.83	30.75	14.25
250						0.74*	0.01	5.00*	30	15
275						0.5	0.05	3.47	31	14
300	0.14	0.36	0.10	32	7	0.62*	0.02	5.00*	30	15
400	0.29	0.20	0.98	28.72	10.28	0.69*	0.01	3.94*	30.67	14.33
500	0.64*	0.02	0.78	29	10	0.49*	0.05	5.25*	29.85	15.15
750	0.43	0.09	1.46	28.18	10.82	0.44	0.07	5.00*	30	15
1000	0.29	0.20	2.83	27	12	0.42	0.08	8.89*	28	17
1250	0.25	0.19	1.64	28	11	0.11	0.35	8.45*	28.20	16.80
1500	0.29	0.20	1.17	10.5	28.5	0.22	0.23	6.81*	16	29

Note. τ: Kendall's tau; χ2: Chi-squared statistic; U80: goal count under the 80th percentile of Success-Scores; O80: goal count above the 80th percentile of Success-Scores; * indicates statistical significance at the level of $p < 0.05$; all values are rounded to two decimal places.

The results of the team comparisons can be found in Tables 5a and 5b for the 16-meter-datasets and in Tables 6a and 6b for the 30-meter-datasets. Due to large sample sizes all differences between Team A and Team B as well as their respective opponents in the Success-Scores across all interval lengths for both 16 and 30 meters were found to be statistically significant. Statistically significantly higher Success-Scores were found for Team B than Team A and conversely, statistically significantly higher Success-Scores were found for Team A's opponents than for Team B's opponents. However, the effect sizes differ depending on the interval length and the area under investigation. The difference between Team A and B was minimal at an interval length of 10 seconds for both 16 and 30 meters. The difference increased with the interval length for both areas. This rise seems to plateau around 1000 seconds for 16 meters and 750 seconds for 30 meters. When comparing the effects for 16 and 30 meters, it is noticeable that the effects are larger for the 30-meter datasets with short intervals of up to around 300 seconds. With longer intervals, the effects are larger for the 16-meter-datasets. Looking at the differences between the two teams' opponents one can see a mirror image of the previously described effects for the 16-meter-Success-Scores. The differences are minimal for the shortest intervals but increase and plateau in a similar way at almost the same effect size. For the 30-meter-Success-Scores the effects are smaller compared to the 16-meter-scores of opponents (Table 5b) as well as the equivalent analysis for Team A and B (Table 6a). The difference between the opponents increases with the interval length of only up to around 100 seconds. Subsequently, the effect actually decreases with any further increase in interval length.

Table 3: Goal counts for the eight percentile groups (columns) of the 16-meter-datasets.

Interval length (s)	0-12.5	12.5-25	25-37.5	37.5-50	50-62.5	62.5-75	75-87.5	87.5-100
10	3.97	3.97	3.97	3.97	3.97	3.97	3.97	11.19
25	4.27	4.27	4.27	4.27	4.27	4.27	5.35	8.00
50	4.33	4.33	4.33	4.33	4.33	4.34	2.00	11.00
100	3.60	3.60	3.60	3.19	7.60	2.40	6.08	8.92
110	3.71	3.71	3.71	3.86	6.00	3.00	7.00	8.00
120	3.37	3.37	3.37	4.90	4.00	5.00	8.00	7.00
130	2.38	2.38	2.38	6.23	5.63	5.00	7.00	8.00
140	3.44	3.44	3.44	5.68	4.00	5.00	6.00	8.00
150	3.33	3.33	3.33	7.00	1.90	6.10	6.00	8.00
200	2.64	2.64	7.63	3.85	6.23	3.00	5.46	7.54
300	2.64	2.36	8.18	5.82	1.00	9.00	6.00	4.00
400	3.52	3.48	3.33	2.67	7.16	6.84	7.48	4.52
500	3.00	3.36	4.64	4.00	7.16	5.84	5.00	6.00
750	3.99	6.01	3.00	4.37	4.63	3.69	6.31	7.00
1000	4.29	4.00	3.82	6.08	2.81	5.45	4.55	8.00
1250	4.17	4.83	5.00	4.00	2.96	4.04	5.00	9.00
1500	4.76	4.24	6.16	2.84	5.00	2.00	6.65	7.35

Note. All values are rounded to two decimal places.

Table 4: Goal count for the eight percentile groups (columns) of the 30-meter-datasets.

Interval length (s)	0-12.5	12.5-25	25-37.5	37.5-50	50-62.5	62.5-75	75-87.5	87.5-100
10	2.35	2.35	2.35	2.35	2.35	2.35	5.92	25.00
25	2.77	2.77	2.77	2.77	2.77	8.16	13.00	10.00
50	1.69	1.69	1.69	3.92	10.00	9.00	7.00	10.00
100	2.69	2.72	2.59	5.00	8.00	5.00	6.00	13.00
150	2.56	2.44	2.00	6.00	3.00	16.00	5.00	8.00
200	1.00	3.00	3.00	5.00	9.00	7.00	11.00	6.00
225	1.56	2.44	4.00	5.00	9.00	5.00	8.00	10.00
250	1.00	3.00	4.00	7.00	9.00	5.00	7.00	9.00
275	2.00	1.00	2.31	10.51	8.92	4.27	6.26	9.74
300	2.00	2.00	4.00	6.00	10.23	3.77	8.00	9.00
400	2.00	2.00	6.00	6.78	6.53	4.35	7.34	10.00
500	2.00	5.00	4.00	8.00	5.00	2.00	8.00	11.00
750	1.00	8.00	6.00	3.00	3.00	4.00	10.00	10.00
1000	4.00	7.00	3.00	4.00	5.00	4.00	6.00	12.00
1250	4.00	7.69	4.31	4.00	4.00	3.00	6.00	12.00
1500	4.00	8.00	4.00	6.00	2.00	5.00	5.00	11.00

Note. All values are rounded to two decimal places.

Table 5a: The Mann-Whitney-U-statistics, associated p-value and the resulting effect size for the 16-meter-datasets for Team A and B.

Interval length (s)	n_1	n_2	$\tilde{x}_A$	$\tilde{x}_B$	U	p-value	ES
10	22381	22259	0	0	235458859	< .0001	0.4726
25	22379	22258	0	0	219693605.5	< .0001	0.4411
50	22382	22258	0	0	201685883	< .0001	0.4048
100	22384	22258	0.1	2.1	207982848	< .0001	0.4174
110	22383	22261	0	2.5	187975818.5	< .0001	0.3773
120	22387	22260	0.1	2.8	186196683.5	< .0001	0.3736
130	22387	22260	0.2	2.9	183172009	< .0001	0.3676
140	22389	22264	0.3	2.9	181618274.5	< .0001	0.3644
150	22389	22260	0.5	3	179557768	< .0001	0.3603
200	22387	22264	1.1	3.6	171941100	< .0001	0.3450
300	22390	22264	1.5	3.8	162722990	< .0001	0.3264
400	22390	22264	1.8	3.7	155999167.5	< .0001	0.3129
500	22390	22264	1.8	3.9	149445074	< .0001	0.2998
750	22390	22264	1.7	3.8	135122115.5	< .0001	0.2711
1000	22390	22264	1.6	3.6	129954128	< .0001	0.2607
1250	22390	22264	1.6	3.3	132199992	< .0001	0.2652
1500	22390	22264	1.4	3.2	132989335.5	< .0001	0.2668

Note. n_1 and n_2: sample sizes for Team A and B; $\tilde{x}_A$ and $\tilde{x}_B$: median for Team A and B; U: Mann-Whitney-U-statistic; ES: effect size; effect sizes are rounded to 4 decimal places.

Table 5b: The Mann-Whitney-U-statistics, associated p-value and the resulting effect size for the 16-meter-datasets for Team A and B's opponents.

Interval length (s)	n_1	n_2	$\tilde{x}_A$	$\tilde{x}_B$	U	p-value	ES
10	22381	22259	0	0	256496560.5	< .0001	0.5149
25	22379	22258	0	0	264933614.5	< .0001	0.5319
50	22382	22258	0	0	279399692	< .0001	0.5608
100	22384	22258	0.7	0	312760721	< .0001	0.6278
110	22383	22261	0.2	0	299203178.5	< .0001	0.6005
120	22387	22260	0.6	0	303141411.5	< .0001	0.6083
130	22387	22260	0.9	0	307968401.5	< .0001	0.6180
140	22389	22264	1.1	0	310020279	< .0001	0.6219
150	22389	22260	1.3	0	312747914.5	< .0001	0.6275
200	22387	22264	1.7	0.5	314507234.5	< .0001	0.6310
300	22390	22264	2.3	0.75	325308164.5	< .0001	0.6526
400	22390	22264	2.6	1.1	335265538	< .0001	0.6726
500	22390	22264	2.6	1.1	343385795.5	< .0001	0.6889
750	22390	22264	2.5	1	361136907.5	< .0001	0.7245
1000	22390	22264	2.4	1	370368819	< .0001	0.7430
1250	22390	22264	2.3	1.1	370372840	< .0001	0.7430
1500	22390	22264	2.1	1	368054734.5	< .0001	0.7383

Note. n_1 and n_2: sample sizes for Team A and B's opponents; $\tilde{x}_A$ and $\tilde{x}_B$: median for Team A and B's opponents; U: Mann-Whitney-U-statistic; ES: effect size; effect sizes are rounded to 4 decimal places.

Table 6a: The Mann-Whitney-U-statistics, associated p-value and the resulting effect size for the 30-meter-datasets for Team A and B.

Interval length (s)	n_1	n_2	$\tilde{x}_A$	$\tilde{x}_B$	U	p-value	ES
10	22379	22258	0	0	228470397	< .0001	0.4587
25	22382	22261	0	0	211022019.5	< .0001	0.4235
50	22381	22260	1.2	9	200530388.5	< .0001	0.4025
100	22379	22260	6.8	14.6	182372806.5	< .0001	0.3661
150	22385	22264	8.3	15.3	176427698.5	< .0001	0.3540
200	22387	22263	8.8	16.4	167265618.5	< .0001	0.3356
225	22390	22263	9.3	16.3	164460075.5	< .0001	0.3299
250	22390	22264	9.6	16.2	164479886	< .0001	0.3300
275	22389	22264	9.5	16	164664897	< .0001	0.3303
300	22390	22264	9.3	15.8	164520592	< .0001	0.3300
400	22390	22264	9.4	15.5	161016723.5	< .0001	0.3230
500	22390	22264	9.2	15.7	156802454	< .0001	0.3146
750	22390	22264	9.5	15.4	154298128.5	< .0001	0.3095
1000	22390	22264	9.6	14.8	157194257.5	< .0001	0.3153
1250	22390	22264	9.3	14.4	158331322.5	< .0001	0.3176
1500	22390	22264	8.5	13.8	159807602	< .0001	0.3206

Note. n_1 and n_2: sample sizes for Team A and B; $\tilde{x}_A$ and $\tilde{x}_B$: median for Team A and B; U: Mann-Whitney-U-statistic; ES: effect size; effect sizes are rounded to 4 decimal places.

Table 6b: The Mann-Whitney-U-statistics, associated p-value and the resulting effect size for the 30-meter-datasets for Team A and B's opponents.

Interval length (s)	n_1	n_2	$\tilde{x}_A$	$\tilde{x}_B$	U	p-value	ES
10	22379	22258	0	0	266909308.5	< .0001	0.5358
25	22382	22261	0	0	281343537	< .0001	0.5647
50	22381	22260	1.2	9	295791812	< .0001	0.5937
100	22379	22260	6.8	14.6	304471080	< .0001	0.6112
150	22385	22264	8.3	15.3	300419618.5	< .0001	0.6028
200	22387	22263	8.8	16.4	293843607	< .0001	0.5896
225	22390	22263	9.3	16.3	291697191.5	< .0001	0.5852
250	22390	22264	9.6	16.2	290862643	< .0001	0.5835
275	22389	22264	9.5	16	290824134.5	< .0001	0.5834
300	22390	22264	9.3	15.8	290497977	< .0001	0.5828
400	22390	22264	9.4	15.5	287269838	< .0001	0.5763
500	22390	22264	9.2	15.7	287687913.5	< .0001	0.5771
750	22390	22264	9.5	15.4	290099600	< .0001	0.5820
1000	22390	22264	9.6	14.8	288569645.5	< .0001	0.5789
1250	22390	22264	9.3	14.4	287246446	< .0001	0.5762
1500	22390	22264	8.5	13.8	281469719	< .0001	0.5646

Note. n_1 and n_2: sample sizes for Team A and B's opponents; $\tilde{x}_A$ and $\tilde{x}_B$: median for Team A and B's opponents; U: Mann-Whitney-U-statistic; ES: effect size; effect sizes are rounded to 4 decimal places.

Discussion

As expected based on theoretical and methodological considerations beforehand, the interval length, as well as the area used to calculate the Success-Score, both largely affect the resulting Success-Score distribution (Perl & Memmert, 2017; Memmert & Raabe, 2018). Shorter intervals for both 16 and 30 meters result in values that range from zero to almost the maximum (100). The shorter the interval (i.e. fewer data points) the more probable a very high to perfect correlation between space control and ball control becomes. Additionally, with shorter intervals, extreme scores are less likely to be averaged with lower scores. On the flip side, this also results in a relatively large share of Success-Scores equaling zero. Logically, there are seconds where no Effort is produced by the attacking team (i.e. the team has space control of less than 20% and no ball control in the analyzed area).

And the shorter the interval the more likely it is to consist of zeros exclusively. Conversely, with increasing interval length the share of zeros in the Success-Score-distribution decreases, as does the range itself. Perhaps less self-explanatory is the observation that mean and median increase with the interval length up to a certain point where the trend is reversed. For longer intervals, the reduction in the Success-Score range is caused by the increase in the number of data points that are being averaged. The initial increase, however, most likely results from the steep decline in the frequency of zeros which is overcompensating the reduction in the Success-Score range. Logically, the median is smaller than the mean due to the overrepresentation of zeros in all distributions. The Success-Scores showed larger values in mean and maximum across all interval lengths when calculated for the final-30-meter-zone. This reflects the nature of the game. For one, the penalty area is smaller than the 30-meter-zone (which includes the penalty area). Subsequently less occurrences of ball control in that area are to be expected. For another, defending teams are assumed to be less willing to concede space and ball control within their own box owing to the proximity to their goal. Overall, the interval length and the chosen area had the expected effects on the resulting Success-Score and its distribution. These shall now be considered in the interpretation of the results.

Goalscoring

16m

Only six out of 17 datasets for the 16-meter-Success-Score revealed a statistically significant rank correlation between the Success-Score and goalscoring. Of these, four were in the range of 110-140 seconds which were included for additional examination of the distributions' characteristics and are expected to be relatively similar due to the small differences in interval length. In contrast to the large effect sizes of those four, the effect sizes for the datasets based on 100 and 150 seconds were small despite the minimal difference in interval length. In general and in these cases, the effects are clearly dependent on singular goals which on their own can change the resulting statistic drastically (Tables 2 and 3). Considering the influence of chance on goalscoring, the variable results are therefore somewhat consequential (Wunderlich et al., 2021). The shorter intervals are clearly impacted by the large number of zeros which cause multiple groups to consist of zeros exclusively resulting in tied ranks which make an effect less likely. Overall, there appears to be a trend of more goals being scored with higher Success-Scores nonetheless. The chi-squared statistics in contrast to Kendall's Tau should be less affected by the share of zeros since grouping is only conducted based on the 80th percentile. Again, the relative goalscoring frequency is higher with higher Success-Scores across all datasets, yet statistical significance is only found for six of the shorter interval lengths. The results vary considerably across interval lengths but it seems to be clear that the link between goalscoring and the Success-Score is smallest when the Success-Score is based on long intervals. In conclusion, no clear link between the Success-Score in the penalty area and goalscoring can be deducted from the results of this study. This comes as a surprise when considering that the penalty area was hypothesized to be a particularly relevant area for goalscoring based on the established connection to match outcome and goalscoring (Liu et al., 2021; Ruiz-Ruiz et al., 2013) in combination with the strong relationship between goalscoring and the Success-Score based on 300 seconds and 30 meters (Brinkjans et al., 2022). Despite the limited relation found, it appeared to be strongest for interval lengths below 200 seconds. These findings fit with the notion that shorter intervals might reflect short phases and events of immediate attacking danger better, while longer intervals might smooth out these events.

Looking closely at Table 3, one could argue that part of the reason for the weak relation is the fact many goals are scored in the midrange of the respective Success-Score. Consequently, it is questionable whether more space control in the penalty area is always proportionally better. Conceivably, a certain amount of space (in combination with ball control) is needed to create a decent goalscoring opportunity. More space might increase the time available to players but is not necessarily the main determinant for the quality of the opportunity. Looking at the highest Success-Scores one might assume that these in contrast represent the situation of unusual freedom in the opponent's box (more or less free on goal) and therefore show another increase in relative goalscoring frequency. Considering the ambiguity of the results, this just represents one possible explanation. Any interpretation should consider that the Success-Score does not distinguish different areas of the penalty area and hence a certain variability in the relation with goalscoring is inevitable.

30m

The same applies to the 30-meter-zone. Nevertheless, only two out of twelve datasets up to the 500-second-interval did not result in a statistically significant rank correlation between the Success-Score and the number of goals scored. Taking into account that these two showed a medium to large effect that is almost statistically significant, it is reasonable to conclude that there is a strong relationship between goalscoring and the Success-Score in the 30-meter-zone based on short to medium intervals. While this effect decreases in the rank correlation for intervals longer than 500 seconds, it remains strong for the comparison of the relative scoring frequency below and above the 80th percentile of the Success-Score. In fact, chi-squared statistics are found to be largest for the short and long intervals, while the effects are smaller in the midrange (150-400 seconds). This could be considered surprising at least with regard to the long intervals. These were hypothesized to smooth out short events, thereby favoring elevated levels of Success-Scores that span longer time periods (i.e. dominance). The extreme effects found for the two shortest intervals confirm that short intervals largely replicate the events. The data under the 80th percentile is almost entirely made up of zeros in those cases. Zeros are most likely to result from no Efforts occurring during the interval. It follows that a team is very unlikely to score a goal if it controls no space and has no ball control in the last 30 meters ten seconds prior to the time point of interest. However, a certain probability is left due to the ways a goal can be scored regardless of a Success-Score of zero: On the one hand, ball control cannot be determined perfectly with position data only, when players of two teams are in close proximity to the ball. Therefore passes, especially crosses (e.g. free kicks) from outside the 30-meter-zone might result in a (headed) goal without an event of ball control being detected. On the other hand, while rare, negative or zero-correlations between ball and space control can on occasion cause the zero-Success-Score and are more probable within shorter intervals. Nevertheless, the shortest intervals appear to be a strong proxy of goalscoring and might be useful as such but reveal little information about the contextual factors contributing to goalscoring. Overall, interval lengths between 100 and 400 seconds for the 30-meter-zone seem to be the most suitable in the analysis of goalscoring. The advantage of the 30-meter-zone is the more frequent occurrence of Efforts leading to more space for variability and hence for a distinction between goal and no goal. Presumably, the penalty area is still more relevant for goalscoring than the 30-meter-zone but the frequency of zeros prevents effects from being detectable. Only with increasing interval lengths do zeros become less dominant. Theoretically, we would expect the link between the Success-Score and goalscoring to decrease due to the large amount of situation-unrelated datapoints being included in each interval. These two factors appear to have opposing effects. Thus, a sweet spot in goalscoring analysis depends on a good balance between these two factors. Within this sample, no sweet spot was detectable for the Success-Score for the penalty area. For the 30-meter-zone an interval length of 100-400 seconds appears to be the optimum. Because of the methodological limitations, the 30-meter-zone seems to be the better choice when analyzing goalscoring even if purely theoretical consideration based on the nature of the game might suggest otherwise.

Inter-team-differences

16m

Based on the previous findings, such as the evidenced relevance of the penalty area (Liu et al., 2021; Ruiz-Ruiz et al., 2013) and the distinction between Team A and B by the Success-Score based on 300 seconds and 30 meters (Brinkjans et al., 2022), inter-term-differences were expected to be apparent in Success-Score for the penalty area. These expectations were confirmed by this analysis primarily via the observed differences. The significance of the difference is only of secondary relevance as the sample size can be considered the primary driver of the significance in this case. The general trend of an increasing difference between the two Teams, as well as their opponents in longer intervals, can be explained by two main reasons. First, the Mann-Whitney-U-test does not compare means or medians but rather the probability of a value from one group being larger than one from the other (Conroy, 2012). Hence, the large number of zeros in the short intervals leads to a lot of ties and therefore the maximum detectable difference between the teams is smaller. Second, in addition, it could be the case that Success-Scores based on longer intervals are more closely related to team quality, perhaps because they represent dominance more accurately and reduce the influence of one-off-events that might favor teams of lower quality or a less dominant playing style in general. At this point such an effect, if present, cannot be separated from the impact caused by the zeros, and in its entiretiy is in no way proven by these results since only two teams were compared. Nevertheless, the plateau in the inter-team-differences (Table 5a and 5b) coincides quite clearly with the diminishing decrease in the number of zeros (Table 1). This indicates that the increase in inter-team difference is primarily caused by the decreasing number of zeros or at least, that intervals above 1000 seconds are not inherently (apart from the resulting distribution) better to distinguish between teams, than the 1000-second interval, regardless of whatever might be reflected by these differences. Consequently, there is no certainty about which interval length inherently has the strongest link to team identity. At this point, it is important to note that even with a decrease in the number of zeros, the test can only reveal a difference if there is one in the non-zero values. Simultaneously, it should be kept in mind that the zeros do not represent some kind of artifact but rather a meaningful Success-Score which, if present in both teams, reflects an actual non-difference. The crucial point is that the frequency of the Success-Scores of zero depends on the parameters used to construct the Success-Score.

The fact that the differences between the two teams' opponents closely mirror those of the two teams themselves hints at the interactive nature of the Success-Scores, which appears to be the case across all interval lengths. The dynamic of the Success-Score entails that this interdependency is synchronized. That means when one team has a high Success-Score chances are that the opponents have a low Success-Score around that same time. With longer intervals, this synchrony inevitably fades away until the Success-Score ends up representing only an overall mean for both teams' performances in that half. Even if such a score would show differences between teams of different quality, it would still lack the dynamic representation of both teams' interaction which is one of the main strengths of the Success-Score. Therefore, a compromise, a sweet spot, must be found once again between the impact that the frequency of zeros has and the loss of dynamic. Then such a Success-Score for the penalty area, as indicated by this analysis, will be higher for better teams that play more dominantly and can be used for team comparisons.

30m

While there are a lot of similarities regarding inter-team differences between the penalty area and the 30-meter-zone, there are also some important differences. As expected, the Success-Score in the 30-meter-zone differs between the two teams across interval lengths. This is consistent with studies highlighting the importance of the final third (Caicedo-Parada et al., 2020; Tenga et al., 2010) as well as the first validation of the Success-Score (Brinkjans et al., 2022). The difference between Team A and B is minimal though significant for the shortest interval and increases up to an interval length of around 200 seconds. The first major difference is the stagnation of differences above 200 seconds. The earlier start of the plateau can be explained by the share of zeros once again. The share continues shrinking for longer intervals than 200 seconds, but only by a few percentage points, resulting in little increase of the potential maximum difference between Teams A and B. The absence of a relevant increase in the effect size for interval lengths between 200 and 1500 seconds, paired with the slight decrease of the zero-share, could be interpreted as an indication that shorter intervals are inherently better suited to distinguish teams when looking at the 30-meter-zone. However, in consideration of the variability in the data, this interpretation seems a bit far-fetched. Perhaps most interestingly, when the influence of the zero-share is minimized, the differences in the 30-meter-zone are considerably smaller than in the penalty area (cf. Table 5a and 6a). The subsequent conclusion is that the penalty area is more relevant in the distinction of teams. Considering that some defending teams might willingly concede control over some areas within the 30-meter-zone at some times but not in their box, Success-Scores within the penalty area are expected to be harder to achieve. Moreover, there is arguably more variability in the value of subspaces within the last 30 meters than in the penalty area (i.e. more areas of lower attacking value). Consequently, the penalty area can be expected to be more closely linked to team quality. Therefore, the results are in accordance with previous findings indicating the relevance of the penalty area (Liu et al., 2021; Ruiz-Ruiz et al., 2013) even though these studies found a connection to match outcome and goalscoring rather than team quality. Another major difference to the penalty-area data can be found in the comparison of the opponents. In the case of the 30-meter-zone, they do not mirror the results of the comparison between Team A and B. In fact, the overall differences are smaller and peak at an interval length of 100 seconds. For intervals longer than 100 seconds the difference between the two teams decreases slightly. This strengthens the discussed interpretation that the intervals of shorter length (here ca. 100 seconds) might be more closely related to team quality. One could argue that the shorter the intervals are the more frequently a team needs to achieve some sort of Success in the analyzed area to minimize the amount of very small or even zero Success-Scores At this point, the authors consider neither this nor any other theoretical (football-related) or methodological (statistical) explanation for shorter intervals to more accurately represent team quality fully convincing. On the contrary, the opposite was expected as shorter intervals more closely represent individual situations which tend to disappear with longer intervals. Evidently, the results allow no clear conclusion about the optimal interval length for the Success-Score in the 30-meter-zone when it comes to team comparisons.

Overall, the findings indicate that the Success-Score can be used to distinguish between teams. This was indicated by this specific comparison between two teams of different quality. It does not follow that team quality differences are the reason for these differences in Success-Scores. In fact, the differences in Success-Scores could be down to any number of differences between the two teams, most notably playing style (Castellano & Pic, 2019). Rather the result highlight, the vast influence of interval length and area analyzed on the Success-Score and associated analyses. The Success-Score for the penalty area shows larger differences between the two teams but is limited by larger numbers of zero at shorter interval lengths. Considering that it would be ill-advised to discard the interactive nature of the Success-Score by choosing extremely long intervals, interval lengths within the previously suggested limits (Perl & Memmert, 2017) are preferable. When comparing the two areas in this range, one can see that they show similar inter-team differences at an interval length of 300 seconds. For the 30-meter-zone interval lengths between 100 and 300 seconds might be optimal. In contrast, for the penalty area, interval lengths of 300 seconds or more are advisable.

In the development of any KPI, its simplicity is essential and has to be balanced with the inclusion of relevant factors. In the case of the Success-Score, this applies to several factors. Space control is computed without accounting for offside. This means that players who are offside and are therefore not allowed to receive the ball from a teammate still contribute to their teams' controlled spaces. A detailed discussion of the relevance of spaces controlled by players behind the offside line, as these are allowed to receive the ball indirectly, is beyond the scope of this article. An additional segmentation of the investigated area could be justified in light of the fact that subsections in those areas inevitably vary in their value. While such factors would undoubtedly add valuable information to the model, they would increase complexity considerably, making interpretation more complex. Moreover, new limitations would inevitably arise. The value of spaces might differ depending on a team's tactics and so might the value of spaces controlled by players that are offside. Furthermore, space control models incorporating player kinematics could improve the performance of the Success-Score and could be considered as potential improvements in the future (Caetano et al., 2021; Fernandez & Bornn, 2018; Martens et al., 2021; Spearman et al., 2017). When working with the Success-Score, awareness of the limitations resulting from position data is required. As described earlier, ball control cannot be determined accurately if players of both teams are in close proximity to the ball. Furthermore, while more representative than the phase between its origin and execution, the Success-Score prior to a set-piece is not necessarily perfectly representative for the execution of the set-piece. Future investigations of the Success-Score should strive to analyze larger datasets including a more variable selection of teams and a larger number of goals to account for the influence of chance. Moreover, future analyses should cover different leagues and competitions to assess the universatility of the Success-Score. Simultaneously, a larger number of teams should be included, showing different combinations of playing style, team quality and other characteristics to ensure that effects can be attributed accurately to either playing style or quality. Although the term team quality was used in this study frequently, it needs to be emphasized once again, that the two teams compared differ in playing style, success, and a theoretically infinite number of characteristics. Therefore, any differences found between the two teams' Success-Scores cannot be attributed to any one of their characteristics specifically.

Conclusion

In its entirety, this study confirmed that the Success-Score is related to goalscoring. This was true across different interval lengths and for both the penalty area and the 30-meter-zone. As hypothesized, the optimal interval length (sweet spot) depends on the interaction of statistical and theoretical considerations. Longer intervals minimize some of the methodological limitations but disturb the dynamic character of the Success-Score. Inter-team differences are best represented by Success-Scores for interval lengths of 300-500 seconds for the penalty area. If the main interest is goalscoring or attacking prowess, the 30-meter-zone is preferable and an interval length between 100 and 400 seconds is appropriate. If both goalscoring and team-differences are relevant, the results suggest that a Success-Score for the 30-meter-zone with an interval length of about 300 seconds is the best compromise.

References

Alves, D. L., Osiecki, R., Palumbo, D. P., Moiano-Junior, J. V. M., Oneda, G., & Cruz, R. (2019). What variables can differentiate winning and losing teams in the group and final stages of the 2018 FIFA World Cup? *International Journal of Performance Analysis in Sport*, 19(2), 248–257. https://doi.org/10.1080/24748668.2019.1593096

Brinkjans, D., Memmert, D., Imkamp, J., & Perl, J. (2022). Success-Score in Professional Soccer – Validation of a Dynamic Key Performance Indicator Combining Space Control and Ball Control within Goalscoring Opportunities. *International Journal of Computer Science in Sport*, 21(2), 32–42. https://doi.org/doi:10.2478/ijcss-2022-0009

Caetano, F. G., Barbon Junior, S., Torres, R. D. S., Cunha, S. A., Ruffino, P. R. C., Martins, L. E. B., & Moura, F. A. (2021). Football player dominant region determined by a novel model based on instantaneous kinematics variables. *Scientific Reports*, 11(1), 18209. https://doi.org/10.1038/s41598-021-97537-4

Caicedo-Parada, S., Lago-Peñas, C., & Ortega-Toro, E. (2020). Passing Networks and Tactical Action in Football: A Systematic Review. *International Journal of Environmental Research and Public Health*, 17(18), 6649. https://doi.org/10.3390/ijerph17186649

Castellano, J., & Pic, M. (2019). Identification and Preference of Game Styles in LaLiga Associated with Match Outcomes. *International Journal of Environmental Research and Public Health*, 16(24), 5090. https://doi.org/10.3390/ijerph16245090

Collet, C. (2013). The possession game? A comparative analysis of ball retention and team success in European and international football, 2007–2010. *Journal of Sports Sciences*, 31(2), 123–136. https://doi.org/10.1080/02640414.2012.727455

Conroy, R. M. (2012). What Hypotheses do "Nonparametric" Two-Group Tests Actually Test? *The Stata Journal: Promoting Communications on Statistics and Stata*, 12(2), 182–190. https://doi.org/10.1177/1536867X1201200202

Dehesa, R., Vaquera, A., Gonçalves, B., Mateus, N., Gomez-Ruano, M. Á., & Sampaio, J. (2019). Key game indicators in NBA players' performance profiles. *Kinesiology*, 51(1), 92–101. https://doi.org/10.26582/k.51.1.9

Fernandez, J., & Bornn, L. (2018). Wide Open Spaces: A statistical technique for measuring space creation in professional soccer. *MIT Sloan Sports Analytics Conference.*

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.

Fujimura, A., & Sugihara, K. (2005). Geometric analysis and quantitative evaluation of sport teamwork. *Systems and Computers in Japan*, 36(6), 49–58. https://doi.org/10.1002/scj.20254

Gollan, S., Ferrar, K., & Norton, K. (2018). Characterising game styles in the English Premier League using the "moments of play" framework. *International Journal of Performance Analysis in Sport*, 18(6), 998–1009. https://doi.org/10.1080/24748668.2018.1539383

Herold, M., Kempe, M., Bauer, P., & Meyer, T. (2021). Attacking Key Performance Indicators in Soccer: Current Practice and Perceptions from the Elite to Youth Academy Level. *Journal of Sports Science and Medicine*, 20, 158–169. https://doi.org/10.52082/jssm.2021.158

Jones, P. D., James, N., & Mellalieu, S. D. (2004). Possession as a performance indicator in soccer. *International Journal of Performance Analysis in Sport*, 4(1), 98–102. https://doi.org/10.1080/24748668.2004.11868295

Kempe, M., Vogelbein, M., Memmert, D., & Nopp, S. (2014). Possession vs. Direct Play: Evaluating Tactical Behavior in Elite Soccer. *International Journal of Sports Science*, 7.

Kirkwood, B. R., Sterne, J. A. C., & Kirkwood, B. R. (2003). Essential medical statistics (2nd ed). Blackwell Science.

Lago-Peñas, C., Lago-Ballesteros, J., & Rey, E. (2011). Differences in performance indicators between winning and losing teams in the UEFA Champions League. *Journal of Human Kinetics*, 27(2011), 135–146. https://doi.org/10.2478/v10078-011-0011-3

Liu, H., Gomez, M.-Á., Lago-Peñas, C., & Sampaio, J. (2015). Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. *Journal of Sports Sciences*, 33(12), 1205–1213. https://doi.org/10.1080/02640414.2015.1022578

Liu, H., Hopkins, W. G., & Gómez, M.-A. (2016). Modelling relationships between match events and match outcome in elite football. *European Journal of Sport Science*, 16(5), 516–525. https://doi.org/10.1080/17461391.2015.1042527

Liu, H., Yi, Q., Giménez, J.-V., Gómez, M.-A., & Lago-Peñas, C. (2015). Performance profiles of football teams in the UEFA Champions League considering situational efficiency. International *Journal of Performance Analysis in Sport*, 15(1), 371–390. https://doi.org/10.1080/24748668.2015.11868799

Liu, T., Yang, L., Chen, H., & García-de-Alcaraz, A. (2021). Impact of Possession and Player Position on Physical and Technical-Tactical Performance Indicators in the Chinese Football Super League. *Frontiers in Psychology*, 12, 722200. https://doi.org/10.3389/fpsyg.2021.722200

Lord, F., Pyne, D. B., Welvaert, M., & Mara, J. K. (2020). Methods of performance analysis in team invasion sports: A systematic review. *Journal of Sports Sciences*, 38(20), 2338–2349. https://doi.org/10.1080/02640414.2020.1785185

Low, B., Coutinho, D., Gonçalves, B., Rein, R., Memmert, D., & Sampaio, J. (2019). A systematic review of collective tactical behaviours in football using positional data. *Sports Medicine,50*, 343-385.

Mackenzie, R., & Cushion, C. (2013). Performance analysis in football: A critical review and implications for future research. *Journal of Sports Sciences*, 31(6), 639–676. https://doi.org/10.1080/02640414.2012.746720

Mao, L., Peng, Z., Liu, H., & Gómez, M.-A. (2016). Identifying keys to win in the Chinese professional soccer league. *International Journal of Performance Analysis in Sport*, 16(3), 935–947. https://doi.org/10.1080/24748668.2016.11868940

Martens, F., Dick, U., & Brefeld, U. (2021). Space and Control in Soccer. *Frontiers in Sports and Active Living*, 3, 676179. https://doi.org/10.3389/fspor.2021.676179

Memmert, D. (Ed.) (2021). *Match Analysis.* Abingdon: Routledge.

Memmert, D., & Raabe, D. (2018). *Data Analytics in Football. Positional Data Collection, Modelling and Analysis.* Abingdon: Routledge.

Memmert, D., & Rein, R. (2018). Match analysis, Big Data and tactics: Current trends in elite soccer. *Deutsche Zeitschrift Für Sportmedizin*, 2018(03), 65–72. https://doi.org/10.5960/dzsm.2018.322

Memmert, D., Lemmink, K. A. P. M., & Sampaio, J. (2017). Current Approaches to Tactical Performance Analyses in Soccer using Position Data. *Sports Medicine*, *47*(1), 1-10.

Memmert, D., Imkamp, J., & Perl, J. (2021). Flexible defends succeeds creative attacks! – A simulation approach based on position data in professional football. *Journal of Software Engineering and Applications, 14(9). DOI: 10.4236/jsea.2021.149029*

Perl, J., Grunz, A., & Memmert, D. (2013). Tactics Analysis in Soccer – An Advanced Approach. *International Journal of Computer Science in Sport*, 12(1), 33–44.

Perl, J., & Memmert, D. (2011). Net-Based Game Analysis by Means of the Software Tool SOCCER. *International Journal of Computer Science in Sport*, 10(2), 77–84.

Perl, J., & Memmert, D. (2017). A Pilot Study on Offensive Success in Soccer Based on Space and Ball Control – Key Performance Indicators and Key to Understand Game Dynamics. *International Journal of Computer Science in Sport*, 16(1), 65–75. https://doi.org/10.1515/ijcss-2017-0005

Perl, J., & Memmert, D. (2018). Soccer: Process and interaction. In A. Baca & J. Perl, *Modelling and Simulation in Sport and Exercise* (S. 73–94). Routledge.

Petersen, C. J. (2017). Comparison of performance at the 2007 and 2015 Cricket World Cups. International *Journal of Sports Science & Coaching*, 12(3), 404–410. https://doi.org/10.1177/1747954117711338

Phatak, A. A., Mehta, S., Wieland, F.-G., Jamil, M., Connor, M., Bassek, M., & Memmert, D. (2022). Context is key: Normalization as a novel approach to sport specific preprocessing of KPI's for match analysis in soccer. *Scientific Reports*, 12(1117). https://doi.org/10.1038/s41598-022-05089-y

Prematunga, R. K. (2012). Correlational analysis. *Australian Critical Care*, 25(3), 195–199. https://doi.org/10.1016/j.aucc.2012.02.003

Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. *SpringerPlus*, 5(1), 1410. https://doi.org/10.1186/s40064-016-3108-2
Rein, R., Raabe, D., & Memmert, D. (2017). “Which pass is better?” Novel approaches to assess passing effectiveness in elite soccer. *Human Movement Science*, 55, 172–181. https://doi.org/10.1016/j.humov.2017.07.010
Ruiz-Ruiz, C., Fradua, L., Fernández-GarcÍa, Á., & Zubillaga, A. (2013). Analysis of entries into the penalty area as a performance indicator in soccer. *European Journal of Sport Science*, 13(3), 241–248. https://doi.org/10.1080/17461391.2011.606834
Sarmento, H., Marcelino, R., Anguera, M. T., CampaniÇo, J., Matos, N., & LeitÃo, J. C. (2014). Match analysis in football: A systematic review. *Journal of Sports Sciences*, 32(20), 1831–1843. https://doi.org/10.1080/02640414.2014.898852
Schlenger, J., Wunderlich, F., Raabe, D., & Memmert, D. Systematic Analysis of Position-Data-based Key Performance Indicators. *International Journal of Computer Science in Sport*, *22*(1), 80-101.
Spearman, W., Basye, A., Dick, G., Hotovy, R., & Pop, P. (2017). Physics-Based Modeling of Pass Probabilities in Soccer. *MIT Sloan Sports Analytics Conference*.
Taki, T., & Hasegawa, J. (2000). Visualization of dominant region in team games and its application to teamwork analysis. *Proceedings Computer Graphics International 2000*, 227–235. https://doi.org/10.1109/CGI.2000.852338
Tenga, A., Ronglan, L. T., & Bahr, R. (2010). Measuring the effectiveness of offensive match-play in professional soccer. *European Journal of Sport Science*, 10(4), 269–277. https://doi.org/10.1080/17461390903515170
Vogelbein, M., Nopp, S., & Hökelmann, A. (2014). Defensive transition in soccer – are prompt possession regains a measure of success? A quantitative analysis of German Fußball-Bundesliga 2010/2011. *Journal of Sports Sciences*, 32(11), 1076–1083. https://doi.org/10.1080/02640414.2013.879671
Winter, C., & Pfeiffer, M. (2016). Tactical metrics that discriminate winning, drawing and losing teams in UEFA Euro 2012®. *Journal of Sports Sciences*, 34(6), 486–492. https://doi.org/10.1080/02640414.2015.1099714
Wunderlich, F., Seck, A., & Memmert, D. (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), 2322–2337. https://doi.org/10.1080/02640414.2021.1930685
Zhou, C., Lago-Peñas, C., Lorenzo, A., & Gómez, M.-Á. (2021). Long-Term Trend Analysis of Playing Styles in the Chinese Soccer Super League. *Journal of Human Kinetics*, 79(1), 237–247. https://doi.org/10.2478/hukin-2021-0077

Appendix

Link to all relevant scripts for the analysis: https://github.com/DavidB1999/SOCCER22