

Evaluating the Wide Attack

Lori He

1 Analysis Scope & Methods



Data: Tracking data and event data of all games in WSL 2023-2024

Objective: Identify and evaluate actions after a player receives a pass in the Wide Attack Area.

- Wide Attack Area: outside of the box and in the attacking third.

Methods:

- Define the success of the attacking sequence following the initial pass.
- Use game context to evaluate the difficulty of success.
- Evaluate and rank players in different circumstances

1.1 Define Candidate Sequences and Success Metrics



Identify Candidate Sequences: Use both tracking data and event data to find receive frame and receiver location of each pass. Then extract passes that receiver location is outside of the box and in the attacking third.

- Receive frame: The frame when the distance between the ball and the receiving player first becomes less than 2 meters.
- Start of sequence: The receive frame.
- End of sequence: When the second pass or shot event occurs after the sequence starts.

Define Success: A sequence is considered successful if it meets the following criteria:

- The ball's new location is at least 2 meters closer to the goal compared to its position at the start of the sequence.
- The attacking team maintains continuous possession of the ball until it reaches the specified location.

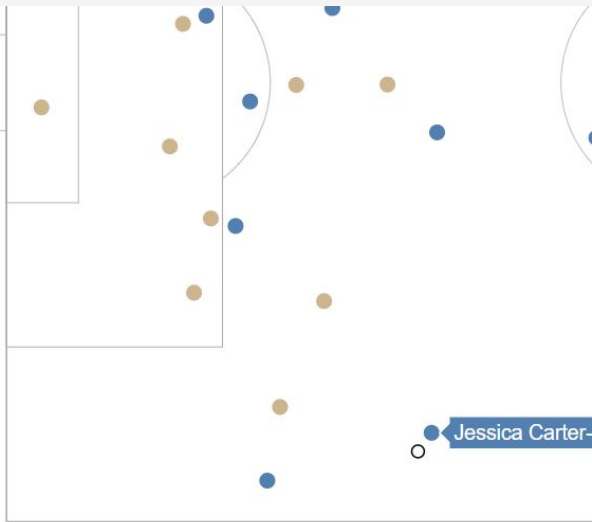
1.2 Example of a Sequence



Here is an example of a sequence which meets the criteria of success at the 2nd event.

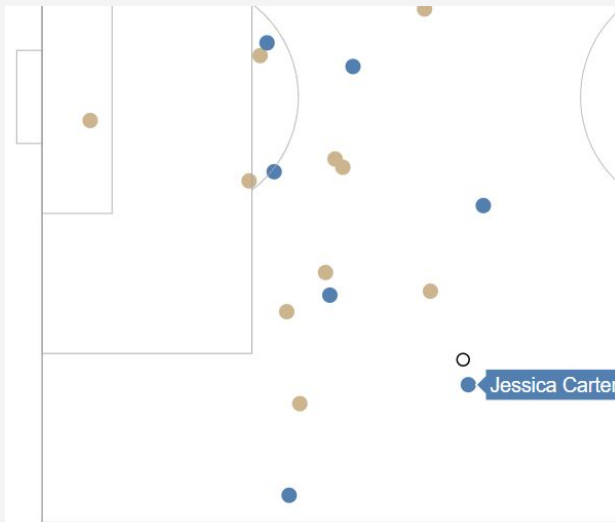
Start of sequence

Ball to goal distance: 42.04 m



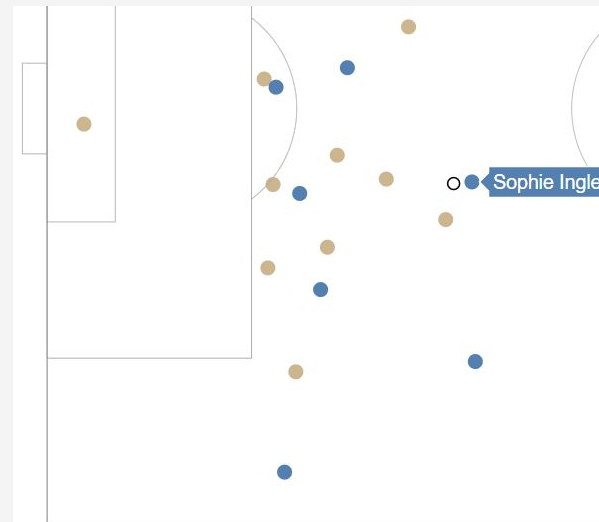
1st event

Ball to goal distance: 40.38 m



2nd event

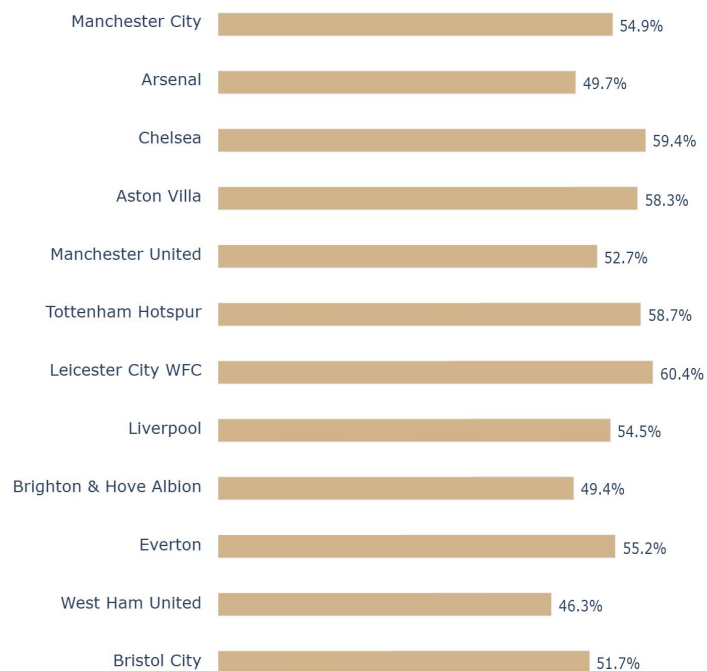
Ball to goal distance: 34.68 m - **Success**



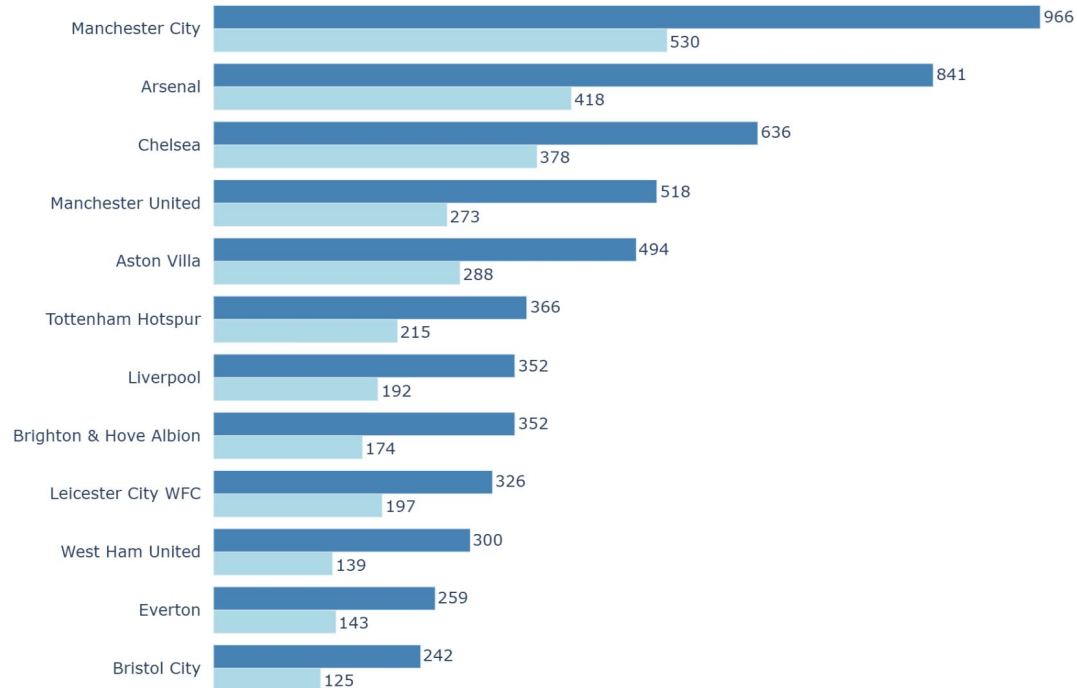
1.3 Team Performance



Manchester City and Arsenal lead in the total number of successful sequences. However, some teams with much lower counts achieve impressive success rates—for instance, Leicester City succeeded in 60.4% of all their wide attack sequences, top of all teams..



■ Success Rate



■ Total Counts

■ Success Counts

2 Context Variables



As the team performance differs, each sequence faces difference attacking difficulties. Here are a few variables considered to have influences on the sequence result.

- **Score difference**

The scoring situation at the start of sequence can affect players' aggressive levels. It may also reflect the skill disparity between the teams.

- **Count of defenders in the box**

With a sufficient number of defenders in the box, the main defender of the sequence has more freedom to leave space behind and press hard, thereby increasing the attacking difficulty.

- **Count of attackers (beyond the attacking point)**

From the 2 meters behind ball location to the goal line, all the attackers within this zone are considered as potential helpers to assist the sequence.

- **Home vs Away**

Playing at a home stadium is often believed to impact a team's performance positively.

- **Sum of Freedom**

The available space where an attacker can move with minimal threat. This is calculated using a Voronoi diagram, as detailed on the next page.

2.1 Sum of Freedom

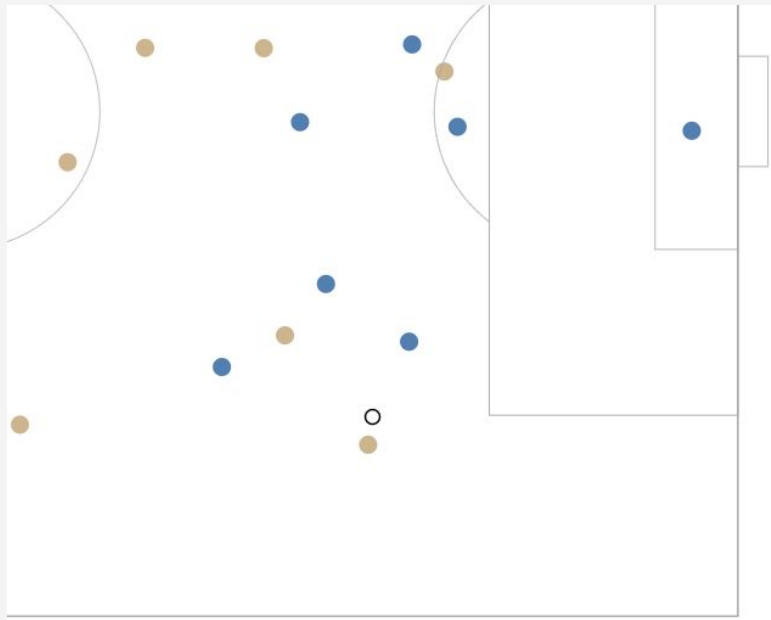
For each sequence, a Voronoi site is generated for the ball and nearby defenders. The site is divided by a line perpendicular to the line connecting the ball and the center of the goal. The Sum of Freedom is defined as the sum of the site areas closer to and farther from the goal, with a heavier weight applied to the area closer to the goal



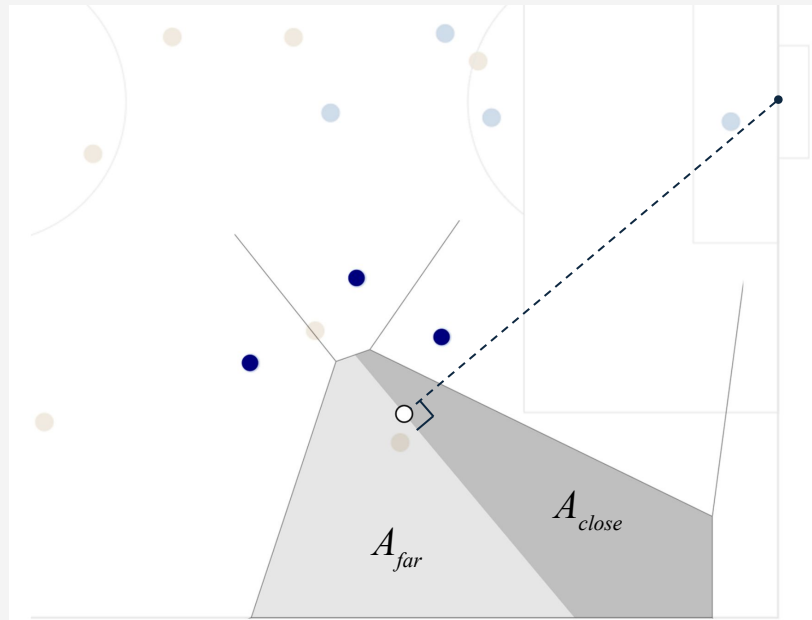
The boundary of the Voronoi diagram was the sidelines, the goal lines, and 20 m away from the ball to four directions.

$$SoF = \log(A_{far} + 1) + A_{close}$$

Example 1 start of sequence



SoF = 166.9 m² Result: Success



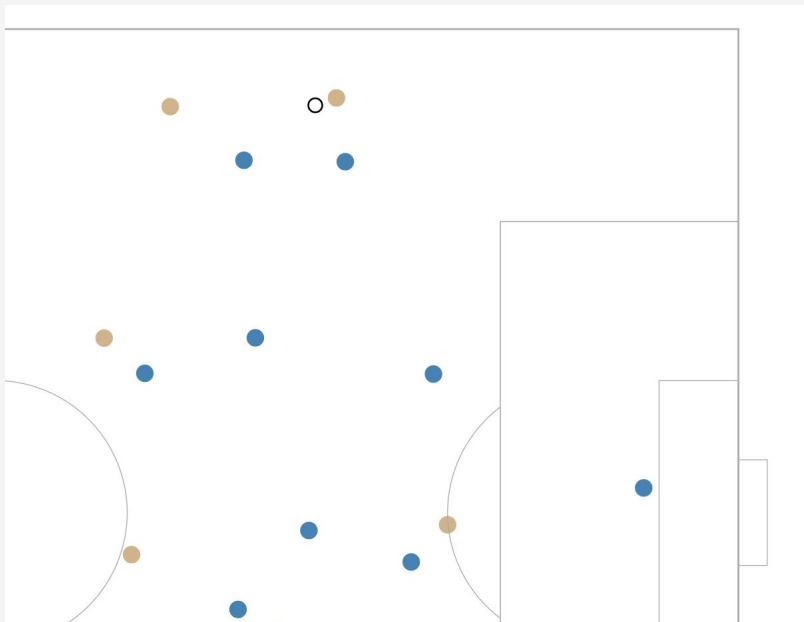
2.1 Sum of Freedom

For each sequence, a Voronoi site is generated for the ball and nearby defenders. The site is divided by a line perpendicular to the line connecting the ball and the center of the goal. The Sum of Freedom is defined as the sum of the site areas closer to and farther from the goal, with a heavier weight applied to the area closer to the goal

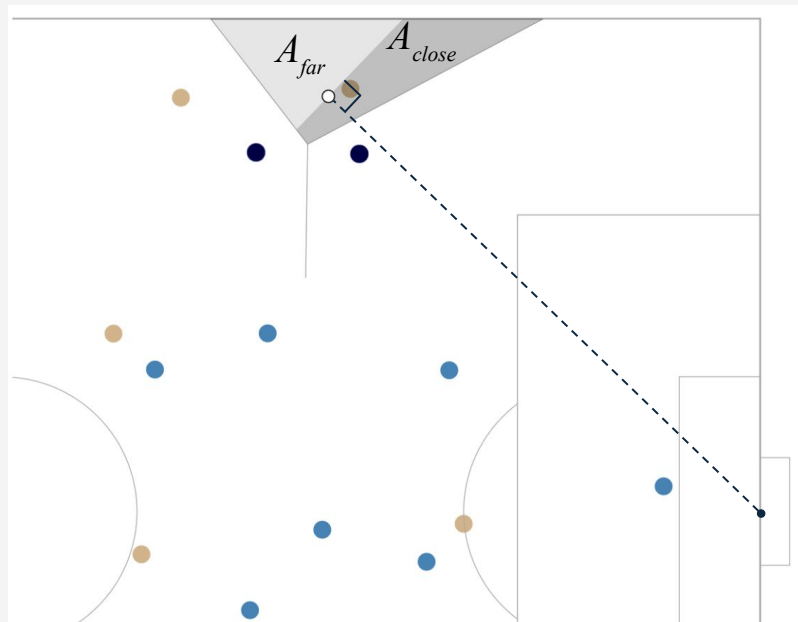
The boundary of the Voronoi diagram was the sidelines, the goal lines, and 20 m away from the ball to four directions.

$$SoF = \log(A_{far} + 1) + A_{close}$$

Example 2 start of sequence

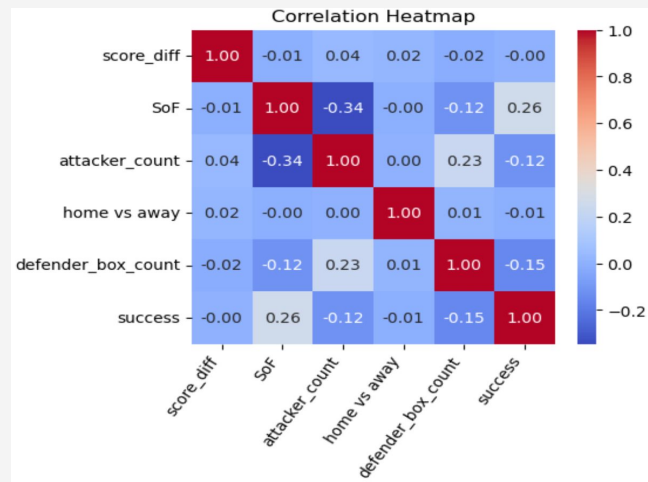
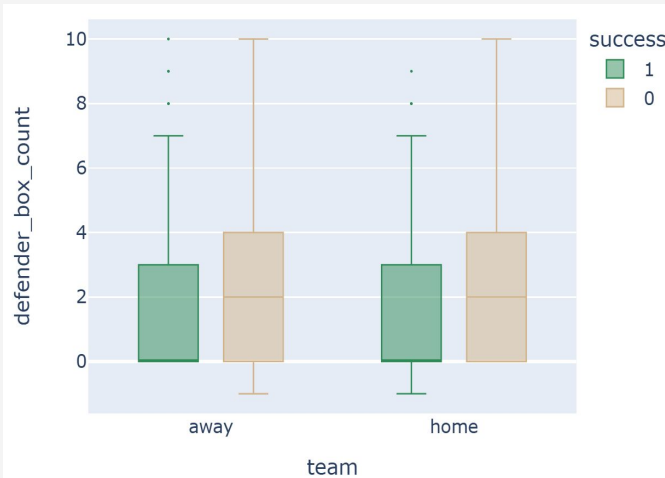
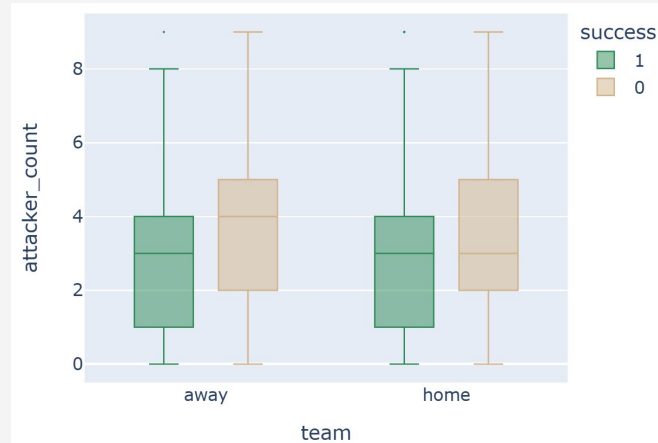
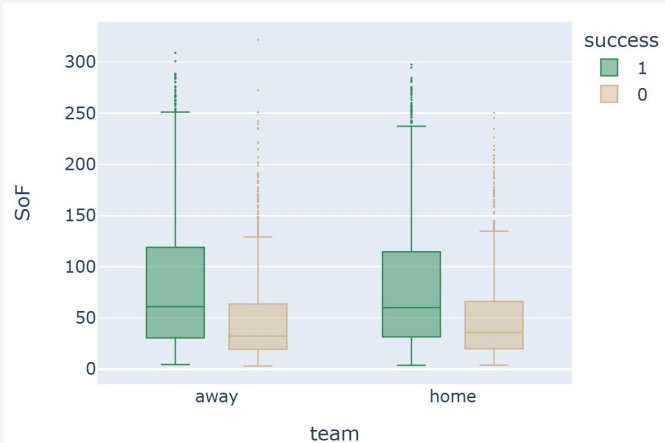


SoF = 51.3 m² Result: non-success



2.2 Feature Validation

A few feature exploration reveals the significance of SoF, defender_box_count, and attacker_count to the sequence result.



3 Models



Using the variables mentioned as features, a few classification models were trained to reveal how much variance of the sequence result can be explained by contextual influences.

In the report on the right, label 1 represents success, label 0 represent unsuccess.

3392 randomly selected sequences were left out before the model training and was used for player evaluation in the next section..

Model: RandomForest					
Accuracy: 0.6438					
Classification Report:					
	precision	recall	f1-score	support	
0	0.59	0.67	0.63	203	
1	0.70	0.63	0.66	249	
accuracy			0.64	452	
macro avg	0.64	0.65	0.64	452	
weighted avg	0.65	0.64	0.64	452	
Confusion Matrix:					
[[135 68]					
[93 156]]					

Model: LogisticRegression					
Accuracy: 0.6283					
Classification Report:					
	precision	recall	f1-score	support	
0	0.56	0.75	0.65	203	
1	0.72	0.53	0.61	249	
accuracy			0.63	452	
macro avg	0.64	0.64	0.63	452	
weighted avg	0.65	0.63	0.63	452	
Confusion Matrix:					
[[153 50]					
[118 131]]					

Model: GradientBoosting					
Accuracy: 0.6372					
Classification Report:					
	precision	recall	f1-score	support	
0	0.57	0.78	0.66	203	
1	0.75	0.52	0.61	249	
accuracy			0.64	452	
macro avg	0.66	0.65	0.64	452	
weighted avg	0.67	0.64	0.63	452	
Confusion Matrix:					
[[159 44]					
[120 129]]					

Model: XGBoost					
Accuracy: 0.6195					
Classification Report:					
	precision	recall	f1-score	support	
0	0.56	0.69	0.62	203	
1	0.69	0.56	0.62	249	
accuracy			0.62	452	
macro avg	0.63	0.63	0.62	452	
weighted avg	0.63	0.62	0.62	452	
Confusion Matrix:					
[[140 63]					
[109 140]]					

4 Player Evaluation



The variance that were not explained by contextual influences should be heavily influenced by the players' capabilities:

- Outstanding attackers would succeed on scenarios which the model considered to be difficult to succeed.
- Outstanding defenders would prevent an attacking success on scenarios considered to be easy to succeed.

Among the four models, the **GradientBoosting** model was chosen to evaluate attacker, because of its better performance on label 0 (unsuccess), the **RandomForest** model chosen to evaluate defender, because of its better performance on label 1 (success).

```
Model: GradientBoosting
Accuracy: 0.6372
Classification Report:
              precision    recall  f1-score   support

     0           0.57       0.78      0.66       203
     1           0.75       0.52      0.61       249

 accuracy                   0.64       452
 macro avg           0.66       0.65      0.64       452
 weighted avg        0.67       0.64      0.63       452

Confusion Matrix:
[[159  44]
 [120 129]]
```

```
Model: RandomForest
Accuracy: 0.6438
Classification Report:
              precision    recall  f1-score   support

     0           0.59       0.67      0.63       203
     1           0.70       0.63      0.66       249

 accuracy                   0.64       452
 macro avg           0.64       0.65      0.64       452
 weighted avg        0.65       0.64      0.64       452

Confusion Matrix:
[[135  68]
 [ 93 156]]
```

4.1 Attacker Evaluation

Among attacker who have faced more than 15 difficult scenarios from the evaluation set, Adriana Leon, Mary Boio Fowler, and Kirsty Hanson ranked high for the percentage of difficult situation they succeeded. However, Chloe Maggie Kelly, Caitlin Jade Foord, and Lauren May Hemp made no less contribution with the total count of difficult situation they succeeded.

recipient_name	success_difficult_ratio	success_difficult	difficult_count
Adriana Leon	0.73	16	22
Mary Boio Fowler	0.63	22	35
Kirsty Hanson	0.61	17	28
Ella Toone	0.59	24	41
Chloe Maggie Kelly	0.57	53	93
Cloe Lacasse	0.55	12	22
Caitlin Jade Foord	0.54	35	65
Lauren May Hemp	0.52	27	52
Sarah Emma Mayling	0.52	14	27
Niamh Louise Charles	0.51	21	41
Lisa Catherine Evans	0.50	11	22
Lauren James	0.50	13	26
Kenza Dali	0.48	15	31
Leila Ouahabi El Ouahabi	0.47	20	43
Grace Clinton	0.43	9	21
Bethany Jane Mead	0.41	18	44



4.1 Defender Evaluation

Defender evaluation was focused on the 'paired players' who were the closest to the ball receiver at the start of sequence.

In defender evaluation, the count of defending an easy to attack situation matters more. Ella Powell's contributions on successfully defending easy attack, facing easy attack, and creating difficult situation to attackers were all on the top tier. Niamh Charles stood out on the ratio of successfully defending easy attack. Heather Payne achieved a balance between the count and the ratio.

paired_player	unsuccess_easy_ratio	unsucceed_easy	easy_count	difficult_count
Ella Mae Florence Powell	0.34	15	44	46
Heather Payne	0.31	10	32	44
Niamh Louise Charles	0.47	9	19	27
Sarah Emma Mayling	0.31	8	26	45
Mayumi Pacheco	0.37	7	19	38
Jamie-Lee Napier	0.28	7	25	57
Ashleigh Neville	0.39	7	18	24
Kerstin Yasmijn Casparij	0.41	7	17	11
Aurora Galli	0.40	6	15	18
Leila Ouahabi El Ouahabi	0.46	6	13	13
Lisa Catherine Evans	0.43	6	14	19
Emma Koivisto	0.24	5	21	33
Kirsty Smith	0.23	5	22	24
Catherine Joan Bott	0.42	5	12	18
Bethany Jane Mead	0.50	5	10	16



THANK YOU

to the ASI and SkillCorner for making this project possible,
and to Tal Levy for the review and feedback.

CONTACT@SKILLCORNER.COM

