# Water Polo Analytics

**Evan Kim, Lucas Novak, Kevin Shen**
evankim@mit.edu, ldnovak@mit.edu, kevshen@mit.edu

June 4th, 2019

# Contents

# 1 Abstract

Although analytics have been used to improve the many different sports, for large part water polo remains untouched. In this report we have watched all 12 games from the Final Eight 2018 European Champion's League Tournament and recorded data from 222 6-on-5 possessions. While this data is slightly limited, our analysis reveals that the 5 position is the weakest, teams should rotate to the right, ending in a 3-3 is very effective way to score, and a skip shot is usually better than a normal shot.

# 2 Introduction

Data analytics has been changing the way sports on all levels have been viewed and played. Despite this, water polo has lagged behind in the use of data analytics to gain a competitive edge. There are only a few papers that attempt to use data science to break down water polo (such as [5]) and USA water polo has a small water polo resource page [6]. However, this is not nearly enough to help water polo coaches on any level improve their game. In this paper we hope to help show the feasibility of water polo analytics by using basic statistical methods to break down 6-on-5 possessions and highlight how this analytical approach can be used to help water polo teams make decisions based on data.

We have chosen to focus on 6-on-5 possessions as the team that scores more goals during 6-on-5 is more likely to win the game [4]. This is also consistent with our data (see Section 3.1). Additionally, 6-on-5 possessions typically have players start in the same position and don't have a large amount of player movement, which makes it easier to record information and less complicated to perform analysis. If we can highlight what leads to more successful 6-on-5 possessions we can hopefully have a stronger impact on helping water polo teams become more successful. First we defined what statistics we thought would be impact the outcome of a 6-on-5. Then we decided to use data from the Final Eight 2018 Champion's League Tournament and manually collected data on each 6-on-5. Afterwards, we applied basic statistical methods to the data.

## 2.1 Data Collection

We choose to collect data from the Final Eight 2018 Champion's League tournament as we wanted to get recent data from the highest level of play possible. We watched the 12 games through Edoardo Osti's Youtube channel [1] and recorded 222 possessions of 6-on-5. This video watching was very time consuming. To speed this up, after 6 games we began to not record player numbers, as finding those with the camera angles provided were both time consuming and unreliable. Additionally we noticed that the player numbers did not add much to our data analysis as they were team specific. For a detailed discussion on the faults of our data collection see Section 2.2.

### 2.1.1 Recording Data

We defined 22 data fields about 6-on-5 possessions to analyze (e.g., was a goal scored and shooter/assisting player position) and an additional 5 data fields to help give identifying information about the play (e.g., team on offense and shooter/assisting player cap number). There is nothing of note to discuss about the identifying information besides we stopped keeping track of the shooter/assisting player cap number after 8 games as it did not add to our analysis and took an inordinate amount of time to determine due to the camera angles. Below we go into a quick discussion about the different fields we collected that we later analyze. For more specific descriptions of each data field see the Appendix. If you want to see this data please see the our GitHub [2].

### 2.1.2 General

A few of the data fields we collect are based on a standard box-score, whether the position lead to a goal, who shot, who assisted, did it lead to a penalty shot, and was the ball turned over. These fields help us analyze whether the 6-on-5 lead to a goal and help identify where on the field the shot and assist came from. We also add a information about the hand of the shooter and assisting player. Additionally, we wanted to include information about how the 6-on-5 possession started, so we recorded if the possession came from the front or back-court and if the possession started from a timeout.

#### 2.1.3 Ball Movement

We wanted a way to record how teams moved the ball during a 6-on-5 as we believed that this is an important factor for a successful 6-on-5 possession. To try to turn ball movement into a data field, we recorded the number of passes and the positions where the ball was passed to. We also combined this data with the time each possession took to get an idea of how fast each team was passing the ball. The number of passes may be skewed up in some cases: if two players are setup early, they tend to pass quickly back and forth while they wait for their teammates to swim to their positions. The data we collected also makes it hard to connect ball movement to player movement. In the future we may try to add a metric that combines player and ball movement to make a combined analysis possible.

#### 2.1.4 Player Movement

Another factor we felt was important to the success of a 6-on-5 was how players moved during a 6-on-5 possession. To capture this we kept track of the rotations that occurred during a possession and whether the possession ended in a 3-3 offensive formation. Players rotate to a general area rather than the same exact place, making recording rotations ambiguous in a few cases.

#### 2.1.5 Shot Type

Where the ball was shot and whether is was blocked or not is very clearly related to the outcome of scoring during a 6-on-5. To measure this, we segmented both the height and the width of the cage into 3 sections, (low, middle, and high) and (left, middle, right) respectively. This created a 3x3 rectangular segmentation of the cage. We also measured if the shot was blocked by the goalie or the field player.

### 2.2 Issues With Data

As mentioned briefly above there are some issues with our data. The most important is defining a player's position. This problem is two-fold. First, position numbers do not translate well between a 4-2 and 3-3 offensive formations. This means that if we were tracking what position a ball was and the offense rotated from one formation to another, we no longer can effectively track where the ball was only using position numbers. Second, position numbers are not very effective at saying where a player or the ball actually is. Position numbers describe a zone rather than an exact location. The location of 1 position in a standard 4-2 is several meters from the position during a 1-in formation and this difference has large ramifications for an offensive possession. A slightly smaller version of this problem also exists with variations in rotations themselves as two players moving 4-center on different possessions can be in slightly different locations. The overall implication of these two issues is that looking at the data does not provide the ability to perfectly understand how the ball moved and to a lesser extent obscures how people actually rotated. A potential way to fix this would to redefine positions to include more specific locations, such as a grid over the front-court or smaller positional zones.

Another problem we experienced was due to the camera angle. The camera being low and to the side made it more difficult to accurately collect data on rotations and shot position and slowed down the data collection process. Additionally, the low, side camera makes the positional problem described above worse by further obscuring where players move (e.g., it is hard to tell if a 4-center is exactly on the middle of the cage or slightly closer to a post). This would also make the solutions to the positional problem we proposed much less effective. If we had a birds eye view of the pool we could much better view player position and a camera along the center of the field could help us know exactly where a player shot a ball.

Another issue worth mentioning is that 3 of us watched games and recorded data individually. While this sped up the process, it does mean that the data each of took will not be 100 percent consistent with each other. We did take steps to mitigate this by coming up with specific rules on how to record each data field (see Appendix) and it is hard to misinterpret straight forwards metrics (e.g., the position of a shooter or assisting player).

## 3 Analysis

We looked at our data from a variety of different perspectives. By doing so we hoped to see the frequency at which different events occurred at the professional level (e.g., how often the 1 shot the ball) and see what actions led the the highest percentage shots (i.e., what led to teams scoring more). Despite the number of games, some factors tended

to have less data and thus the conclusions from the analysis should be taken more lightly. Also note that the data is from the highest level of water polo play and conclusions may not translate perfectly to other levels.

Some of our analysis is based around scoring percentage and some around conversion rate. To clarify, scoring percentage is the percentage of times a shot goes in. Conversion rate is the percentage of 6-on-5s that result in a goal.

For reference the average 6-on-5 conversion rate was 36.94 percent. To see our code please reference our Jupyter Notebook [3]

## 3.1 6-on-5 Percentage and Winning

We wanted to corroborate the results from [4] so we decided to compare the 6-on-5 percentages of winning and losing teams. We found that our results are consistent. The team with a better 6-on-5 conversion percentage won 10 out of 12 times (83.33 percent). Below, seen in Table 1, we also compared the overall conversion rates of winning and losing teams. The winning team had about at 10 percent higher conversion percentage. This was also found to be statistically significant as using a one-sided z-test the p_value was less than 10 percent. We used a one-sided z-test as we were confident that the conversion percentage of winning teams would be higher than losing teams. To see a more detailed discussion on what a z-test is and why is applicable to this data see Section 3.5.

| **Team** | **Goals** | **Attempts** | **Conversion Rate** |
|---|---|---|---|
| **Winning** | 48 | 112 | 42.86 |
| **Losing** | 34 | 110 | 30.91 |

p_value = 0.0326

Table 1: 6-on-5 Conversion Rate Comparison of Winning and Losing Teams

## 3.2 Position Analysis

We compared scoring percentages of shooters and assisting players from each position to find what positions lead to the most efficient shots. For shooters, we also looked at the scoring percentage based on where on the cage they shot the ball. Scoring percentage data based on shot location was not transferred to assisting players, as we did not feel the correlation would be strong enough.

### 3.2.1 Shooters

For each position, we found the number of goals scored, number of shots attempted, the scoring percentage, and broke down the shots that did not go in (blocked by goalie, blocked by a field player, and missed cage completely). The results of this can be seen in Table 2. Both the 1 and 6 took the most shots and had the highest scoring percentage. The position with the worst scoring percentage was the 5. The posts (2, 3) also had a similar scoring percentages, suggesting a shot from either post is worth about the same, although there is some difference in why they their shot blocked or missed.

Below we did analysis on where on cage each shooter shot the ball. Specifically, we divided the cage up into 9 sections based on whether the height of the ball was low, in the middle, or high on the cage and if the ball was on the left, right, or in the center of the cage (based on the shooter's perspective). Below in Tables 3, 4, 5, 6, 7, 8 we added the (goals, attempts, and scoring percentage) to each of these sections. We also included the number of shots that we not on the cage (see Section 3.2.1.7 for more detail). While informative, none of the 9 sections of the cage for any position have enough data for us to feel comfortable about knowing the true percentage of scoring in the different sections of the cage. Additionally, a lot of these shots lack important context (e.g., 1 receiving the ball from 6 may lead to a more open left side of the cage than a pass from 4). That being said, the distribution of shots over the course of 12 games is still indicative of the general shot selection taken by players and opportunities that their movement opened up.

| Position | Goals | Attempts | % Scored | % Blocked By Goalie | % Shot Blocked | % Missed |
|---|---|---|---|---|---|---|
| **1** | 23 | 53 | 43.40 | 33.96 | 7.55 | 15.09 |
| **2** | 14 | 35 | 40.00 | 42.86 | 2.86 | 11.43 |
| **3** | 9 | 22 | 40.91 | 27.27 | 9.09 | 18.18 |
| **4** | 11 | 31 | 35.48 | 25.81 | 12.90 | 25.81 |
| **5** | 7 | 26 | 26.92 | 15.38 | 7.69 | 50.00 |
| **6** | 18 | 36 | 50.00 | 25.00 | 8.33 | 16.67 |

Table 2: Scoring Rates and Block/Miss Rates For Each Position

#### 3.2.1.1 Shoots From 1

| Not on Cage: 5 | | **Left** | **Center** | **Right** | | Not on Cage: 5 |
|---|---|---|---|---|---|---|
| | **High** | (4, 12, 33.3) | (2, 2, 100.0) | (3, 5, 60.0) | **High** | |
| | **Middle** | (3, 5, 60.0) | (0, 3, 0.0) | (0, 2, 0.0) | **Middle** | |
| | **Low** | (7, 10, 70.0) | (0, 1, 0.0) | (3, 5, 60.0) | **Low** | |

Table 3: Position of Shots on Cage From 1

#### 3.2.1.2 Shoots From 2

| Not on Cage: 5 | | **Left** | **Center** | **Right** | | Not on Cage: 5 |
|---|---|---|---|---|---|---|
| | **High** | (6, 9, 66.7) | (0, 5, 0,0) | (2, 3, 66.7) | **High** | |
| | **Middle** | (1,2,50.0) | (0, 0, NA) | (0, 1, 0.0) | **Middle** | |
| | **Low** | (4, 4, 100.0) | (0, 2, 0.0) | (1, 1, 100.0) | **Low** | |

Table 4: Position of Shots on Cage From 2

#### 3.2.1.3 Shoots From 3

| Not on Cage: 3 | | **Left** | **Center** | **Right** | | Not on Cage: 3 |
|---|---|---|---|---|---|---|
| | **High** | (1, 1, 100.0) | (1, 1, 100.0) | (2, 4, 50.0) | **High** | |
| | **Middle** | (0, 0, NA) | (1, 1, 100.0) | (1, 3, 33.3) | **Middle** | |
| | **Low** | (2, 2, 100.0) | (0, 1, 0.0) | (1, 4, 25.0) | **Low** | |

Table 5: Position of Shots on Cage From 3

#### 3.2.1.4 Shoots From 4

| Not on Cage: 2 | | **Left** | **Center** | **Right** | | Not on Cage: 2 |
|---|---|---|---|---|---|---|
| | **High** | (1, 1, 100.0) | (1, 1, 100.0) | (1, 3, 33,3) | **High** | |
| | **Middle** | (0, 0, NA) | (2, 4, 50.0) | (0, 3, 0.0) | **Middle** | |
| | **Low** | (3, 4, 75.0) | (0, 3, 0.0) | (2, 6, 33.3) | **Low** | |

Table 6: Position of Shots on Cage From 4

#### 3.2.1.5 Shoots From 5

Not on Cage: 4

| | Left | Center | Right | |
|---|---|---|---|---|
| **High** | (0, 4, 0.0) | (2, 2, 100.0) | (2, 4, 50.0) | **High** |
| **Middle** | (0, 2, 0.0) | (0, 0, NA) | (0, 0, NA) | **Middle** |
| **Low** | (1, 2, 50.0) | (1, 1, 100.0) | (1, 5, 20.0) | **Low** |

Not on Cage: 4

Table 7: Position of Shots on Cage From 5

#### 3.2.1.6 Shoots From 6

Not on Cage: 3

| | Left | Center | Right | |
|---|---|---|---|---|
| **High** | (2, 5, 40.0) | (1, 1, 100.0) | (2, 3, 66.7) | **High** |
| **Middle** | (2, 3, 66.7) | (1, 2, 50.0) | (1, 2, 50.0) | **Middle** |
| **Low** | (3, 4, 75.0) | (0, 0, NA) | (6, 11, 54.5) | **Low** |

Not on Cage: 3

Table 8: Position of Shots on Cage From 6

**3.2.1.7 Not On Cage** The Not on Cage values include all shots that were not cage (above cage, wide of cage, or both). The data includes which of these cases specifically but we choose to give an aggregate of all cases to simplify visual representation.

### 3.2.2 Assisting Players

We wanted to compare the effectiveness of shots when assisted from each position. In order to do this, we collected the number of passes to a shooter, the number of shoots that shooter scored, and then calculated percent of those passes that lead to a goal. The results are in Table 9. While the posts had the highest percentage, there are only a handful of passes from each post and the post only passes in very situational circumstances. The perimeter players had about the same percentage, although passes from 4 seemed to be the best. It is also interesting that there were about half as many passes from 1 than any other perimeter position and this is probably due to how many more shots 1 took than other position.

| Position | Passes That Lead to Goal | Number of Passes | % Passes Led to Goal |
|---|---|---|---|
| **1** | 10 | 27 | 37.04 |
| **2** | 2 | 2 | 100.0 |
| **3** | 4 | 6 | 66.67 |
| **4** | 24 | 58 | 41.28 |
| **5** | 20 | 51 | 39.22 |
| **6** | 21 | 56 | 37.5 |

Table 9: Pass Conversion Passes From Each Position

### 3.2.3 Shooters and Assisting Players

In this section we combined our approach of the previous two sections. For each shooter, assisting player pair, we found the number of shots attempted, goals scored, and then calculated the scoring percentage. The results of this can be seen in Table 10. What is interesting is what positions brought down the scoring/assisting percentages of their pairs. 5 and 6 shooting passes from each other lead to poor scoring rates, bringing down both of their shot and pass conversion rates in both scenarios. We also found the scoring percentage of the 2 and 3 posts were influenced by the assisting player. The 3 post was great at scoring off of passes from 5, average at scoring from passes from 4, bad at scoring passes from 6, and terrible at scoring passes from 1. The 2 post on the other hand was great at scoring passes from 6, average at scoring passes from 1, and bad at scoring passes from 4 and 5. We still need more data to make any definitive statements about exact conversion rates.

| Shooter Position | Assisting Player Position | Goals | Attempts | % Scored |
|---|---|---|---|---|
| **3** | **5** | 4 | 4 | 100.00 |
| **5** | **3** | 1 | 1 | 100.00 |
| **6** | **2** | 2 | 2 | 100.00 |
| **6** | **3** | 2 | 2 | 100.00 |
| **2** | **6** | 6 | 8 | 75.00 |
| **6** | **1** | 6 | 9 | 66.67 |
| **1** | **5** | 5 | 9 | 55.56 |
| **4** | **3** | 1 | 2 | 50.00 |
| **6** | **4** | 5 | 10 | 50.00 |
| **4** | **6** | 3 | 7 | 42.86 |
| **5** | **4** | 3 | 7 | 42.86 |
| **1** | **4** | 11 | 26 | 42.31 |
| **2** | **1** | 2 | 5 | 40.00 |
| **3** | **4** | 2 | 5 | 40.00 |
| **1** | **6** | 7 | 8 | 38.88 |
| **4** | **5** | 6 | 16 | 37.50 |
| **2** | **4** | 3 | 9 | 33.33 |
| **2** | **5** | 3 | 10 | 30.00 |
| **3** | **6** | 2 | 7 | 28.57 |
| **4** | **1** | 1 | 5 | 20.00 |
| **5** | **6** | 3 | 16 | 18.75 |
| **3** | **1** | 1 | 6 | 16.67 |
| **6** | **5** | 2 | 12 | 16.67 |
| **5** | **1** | 0 | 2 | 0.00 |

Table 10: Conversion Rates For Each Shooter/Assisting Player Pair (if a pair isn't here it is because they never took a shot)

## 3.3 Passes and Timing

We wanted to tell if the number of passes and the time each 6-on-5 possession took had any major effect on the shot percentage. As mentioned in Section 2.1.3, the number of passes in some cases may be skewed up. However, we don't believe this will have a dramatic effect on our analysis.

### 3.3.1 Number of Passes

For each possession, we counted the number of passes and then found the number of possession, goals, and conversion rate associated with the number of passes. The results can be seen in Table 11 and Figure 1. There are spikes at 1 pass and at 4 passes and a general trend of increase in conversion rate with the number of passes. The spike at 1 pass is most likely due to quick passes to the player that earned the ejection in front of the goal before the defense has time to set up. The spike at 4 passes is a bit of a mystery, it is most likely due to random variation and a low number of samples. The large jump in conversion rate from 7 to 8 passes is very interesting. This probably is because after enough passes the offense has probably run a rotation or at least passed the ball enough to create an advantage. While the data might suggest the number of passes is proportional to conversion rate, there is not enough data to support this. We do feel confident that number of passes and conversion rate is at least a threshold, where after enough passes the offense will create a higher percentage shot.

| Number of Passes | Goals | Possession | Conversion Rate |
|---|---|---|---|
| **0** | 1 | 5 | 20.00 |
| **1** | 4 | 6 | 66.67 |
| **2** | 4 | 13 | 30.77 |
| **3** | 4 | 15 | 26.67 |
| **4** | 9 | 15 | 60.00 |
| **5** | 5 | 22 | 22.73 |
| **6** | 7 | 26 | 26.92 |
| **7** | 9 | 37 | 24.32 |
| **8** | 9 | 22 | 40.91 |
| **9** | 14 | 31 | 45.16 |
| **10** | 5 | 12 | 41.67 |
| **11** | 5 | 9 | 55.56 |
| **12** | 2 | 4 | 50.00 |
| **13** | 3 | 4 | 75.00 |

Table 11: Number of Passes and Conversion Rate

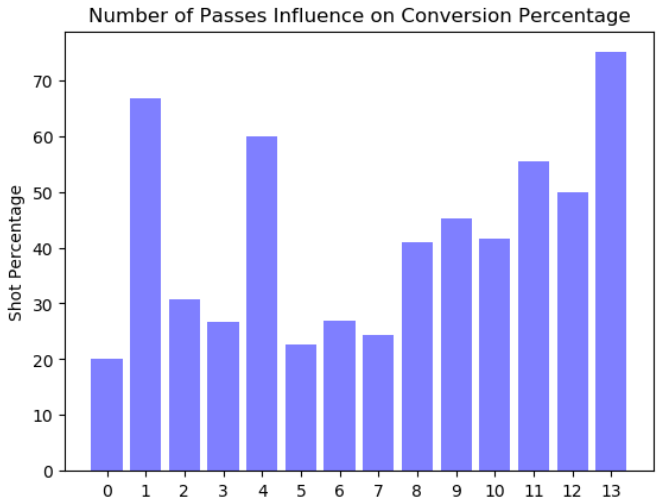


Figure 1

### 3.3.2 Time of Possession

For each possession, we recorded the number of goals and calculated the conversion rate for the time each possessions took. The data can be see in Table 12 and visualized in Figure 3. Because each possession can take up to 30 seconds, data for the time of possession is less than we would have liked and this makes us less confident in any exact percent. However, the results are similar to that of the number of passes. There is a high conversion rate early on, probably for the quick pass or defenses that were not set up effectively. And after enough time the shot percentage jumped up, which is similar to the offense making enough passes. You can also see that after 21 seconds, the conversion rate drops. This suggests that teams should try to attempt a shot at the end of the 6-on-5 before the ejected player can effectively guard someone rather than try to score a goal in the remaining 9 seconds when the ejected player has integrated themselves with the defense.

| Time of Possession (seconds) | Goals | Possessions | Conversion Rate |
|---|---|---|---|
| **1** | 0 | 3 | 0.00 |
| **2** | 2 | 2 | 100.00 |
| **3** | 2 | 2 | 100.00 |
| **4** | 3 | 4 | 75.00 |
| **5** | 3 | 5 | 60.00 |
| **6** | 1 | 4 | 25.00 |
| **7** | 2 | 5 | 40.00 |
| **8** | 1 | 6 | 16.67 |
| **9** | 1 | 4 | 25.00 |
| **10** | 3 | 5 | 60.00 |
| **11** | 2 | 4 | 50.00 |
| **12** | 1 | 5 | 20.00 |
| **13** | 3 | 6 | 50.00 |
| **14** | 2 | 11 | 18.18 |
| **15** | 4 | 9 | 44.44 |
| **16** | 6 | 14 | 42.86 |
| **17** | 3 | 15 | 20.00 |
| **18** | 6 | 16 | 37.50 |
| **19** | 11 | 20 | 55.00 |
| **20** | 8 | 18 | 44.44 |
| **21** | 6 | 14 | 42.86 |
| **22** | 5 | 20 | 25.00 |
| **23** | 2 | 8 | 25.00 |
| **24** | 3 | 10 | 30.00 |
| **25** | 1 | 5 | 20.00 |
| **26** | 1 | 4 | 25.00 |
| **27** | 0 | 1 | 0.00 |
| **28** | 0 | 1 | 0.00 |
| **29** | 0 | 0 | NA |
| **30** | 0 | 1 | 0.00 |

Table 12: Conversion Rate of 6-on-5 Possession by Seconds

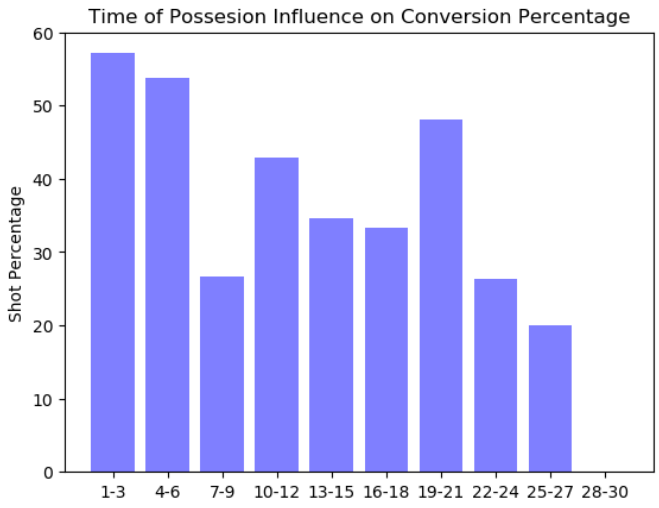


Figure 2

### 3.3.3 Passes per Second

We combined our data from the previous two Sections and calculated the goals and conversion rates for passes per second. The data can be see in Table 12 and visualized in Figure 3. The intervals were created using numpy's histogram package. Similar to the last two Sections, this data supports the effectiveness of the quick pass. A quick pass usually takes 1-3 seconds and that would translate to a .3-.5 passes per second, which is has a higher conversion rate. Similar to the number of passes, there is a large percentage jump around .237 passes per second, which is about 1 pass per 4 seconds. We also have more data around this jump, so we feel more confident that this threshold model exists, although we are not confident with the exact passes per second that causes the jump. Again, we don't feel confident in more passes per second leading to higher conversion rate as the data drops off after .623 passes per second. It is harder to generalize about passes per second because this is the highest level of water polo possible. These players are always a threat with the ball and can pass quickly without turning the ball over. We would need data of other levels of play in order to make that analysis.

| Passes per Second | Goals | Possessions | Conversion Rate |
|---|---|---|---|
| **0.0-0.089** | 1 | 6 | 16.67 |
| **.09-0.178** | 1 | 4 | 25.00 |
| **0.179-0.267** | 7 | 28 | 25.00 |
| **0.268-0.356** | 12 | 48 | 25.00 |
| **0.237-0.444** | 27 | 65 | 41.54 |
| **0.445-0.533** | 16 | 45 | 35.56 |
| **0.534-0.622** | 10 | 15 | 66.67 |
| **0.623-0.711** | 4 | 7 | 57.14 |
| **0.712-0.8** | 2 | 2 | 100.00 |
| **2.0** | 1 | 1 | 100.00 |

Table 13: Conversion Rate Different Passes Per Second

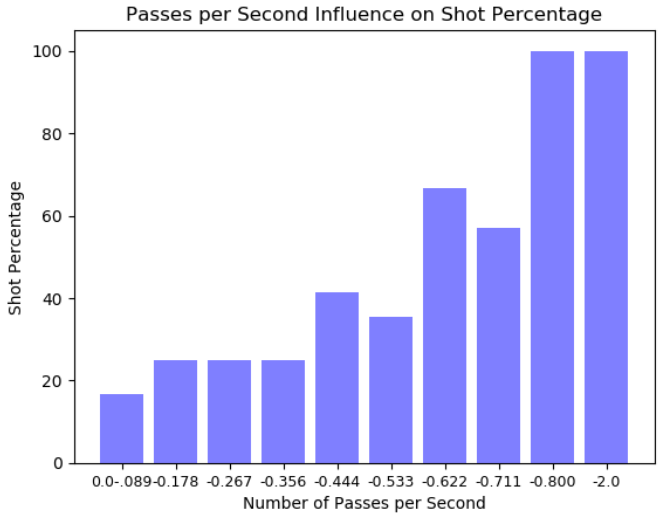


Figure 3

## 3.4 Rotation Types

Rotations are a major component of 6-on-5 offense so determining what rotation types are most effective is crucial to helping us analyze what makes a good 6-on-5. For this section we found the conversion rates of 6-on-5s when different rotation types occurred. The results can be seen in Table 14.

The big thing that jumps out from the data is how much better the conversion rate is when players rotate to

the right (6_in, 4_center, 1_pocket). This makes sense as these players are usually right handed and catch the ball strong-sided, allowing passing to be smoother. Additionally, this is consistent with what we saw with shooting percentages in Section 3.2, as this removes the 5 from the equation and rotations that increased 5's involvement had the worst percentages.

| Rotation Type | Goals | Possessions | Conversion Rate |
|---|---|---|---|
| **6_in+4_center** | 33 | 65 | 50.77 |
| **6_in** | 52 | 113 | 46.02 |
| **1_pocket+4_center** | 15 | 34 | 44.12 |
| **4_center** | 41 | 94 | 43.62 |
| **2_pop** | 14 | 35 | 40.00 |
| **1_pocket** | 19 | 48 | 39.58 |
| **No Rotation** | 10 | 26 | 38.46 |
| **1_in+6_in** | 8 | 26 | 30.77 |
| **3_pop** | 4 | 14 | 28.57 |
| **1_in** | 17 | 60 | 28.33 |
| **1_in+5_center** | 7 | 29 | 24.14 |
| **5_center** | 15 | 65 | 23.08 |
| **6_pocket** | 7 | 38 | 18.42 |
| **6_pocket+5_center** | 4 | 27 | 14.81 |

Table 14: Rotation Type Effect on Conversion Rate

## 3.5 Effects of Other Factors

In this section we compare the affect different metrics have on the scoring percentage or conversion rate. Again for reference the overall 6-on-5 percentage was 36.94 percent.

For these factors we have decided to also test their affect on the outcome using a z-test. This means we want to find the probability that a factor actually contributed to a different shot percentage. For each of the below factors we initially assume that the difference in percentage was caused by natural variation. If the z-test says the probability of different percentages happening is less than 10 percent we will conclude that the factor had an impact on shooting percentage or conversion rate. We choose to have a higher percentage cutoff for the z-test, and thus higher chance at falsely claiming a factor had an impact, as this report is exploratory in nature. We would rather find more factors to explore in the future, and determine some of them irrelevant upon further testing, than find less factors to explore and potentially ignore something relevant.

In order to accurately apply a z-test, we made the following assumptions: the 6-on-5s we observed are representative of random sample of 6-on-5 possessions in general and each 6-on-5 is independent of each other. While all these 6-on-5 possessions were from the same tournament, due to their high level of play we feel like they are an accurate sample of possessions of high level play. And although, like in any sporting event, one possession will have some minor impact on how a team plays for the rest of the game, we feel comfortable considering each 6-on-5 independent. (Note: The other assumptions to use a z-test are also met. There for each factor we have at least 10 goals and there are many more than 20 times the number of 6-on-5s in all water polo games than from our sample).

There is also a distinction between a one-sided and two-sided z-test. Essentially, one-sided tests are for factors we know will definitely increase/decrease the scoring percentage or shooting rate. Two-sided tests are for when we are not confident in whether a factor will increase or decrease the scoring percentage or shooting rate.

### 3.5.1 Dropped Ball

The data for converting a 6-on-5 when a ball is dropped or not dropped is in Table 15. We did a one-sided z-test as we believed dropping the ball has no reason to help the offense, as it kills the flow of the offense and allows to recover. While the percentage was definitely worse when dropping the ball, the p-value isn't low enough for us to confidently conclude this. This is probably because at the highest level of play, an amazing shooter can score at any point, despite any lapses on the offensive end.

| Ball Dropped | Goals | Attempts | Conversion Rate |
|---|---|---|---|
| **No** | 49 | 121 | 40.50 |
| **Yes** | 33 | 101 | 32.67 |

p_value: 0.1146

Table 15

### 3.5.2 Timeout

We compared the conversion rate when a timeout was called or not called and the results can be seen in Table 16. We did a two-sided z-test because we were unsure if timeouts would actually help, they allow the offense to get a play but also allow the defense to set up. While timeouts did improve the 6-on-5 percentage, it was very marginal is a large p-value. In the future we would want to test if timeouts were better and use a one-sided test to do this.

| Timeout Called | Goals | Attempts | Conversion Rate |
|---|---|---|---|
| **No** | 66 | 180 | 36.67 |
| **Yes** | 16 | 42 | 38.10 |

p_value: 0.4314

Table 16

### 3.5.3 Ejected in Front or Back-court

We compared the conversion rate when the ejection occurred in the front-court or back-court and the results can be seen in Table 17. We did a two-sided z-test because we were unsure if front-court or back-court is better. The data showed that back-court ejections were much better for 6-on-5 conversion rate than front-court ejections, but the p-value is not low enough for us to conclude anything. We want to test this further with a one-sided test.

| Front/Back-court | Goals | Attempts | Conversion Rate |
|---|---|---|---|
| **Front-court** | 66 | 184 | 35.87 |
| **Back-court** | 16 | 38 | 42.11 |

p_value: 0.2342

Table 17

### 3.5.4 Ejected in Front or Back-court With Timeouts

We compared the conversion rate when a timeout was called and the ejection occurred in the front-court or back-court and the results can be seen in Table 18. The data showed that with timeouts, back-court ejections were still much better for 6-on-5 conversion rate than front-court ejections, better even than timeouts alone or back-court alone. However, the data is somewhat limited for this analysis and we couldn't even apply a z-test.

| Front/Back-court | Goals | Attempts | Conversion Rate |
|---|---|---|---|
| **Front-court** | 10 | 29 | 34.48 |
| **Back-court** | 6 | 13 | 46.15 |

Table 18

### 3.5.5 End in 3-3

We compared the conversion rates of 6-on-5s that ended in a 3-3 and that didn't end in a 3-3 and the results can be seen in Table 19. We applied a two-sided z-test as we didn't know if that would increase of decrease the conversion percentage. The conversion rate was much better for ending in a 3-3 and the p-value is low enough for us to conclude that ending in a 3-3 makes a difference. This is probably because ending in a 3-3 can create more room for error and cause the defense to be out of position if not guarded correctly (e.g., two people guarding set and leaving a wing open).

| Ended in 3-3 | Goals | Attempts | Conversion Rate |
|---|---|---|---|
| **No** | 69 | 200 | 34.50 |
| **Yes** | 13 | 22 | 59.09 |

p_value: 0.0233

Table 19

### 3.5.6 Left vs Right Hand

For each position, we compared the scoring percentage and assisting rate for players that were left and right handed. The results can be seen in Tables 20 and 21. For both we found very little difference between each position and when we did there is not enough data to back it up. We applied a two-sided z-test to the combined 5 and 6 percentages for both shooting and assisting. For more detail on why see Section 3.5.6.3.

| Position | Hand | Goals | Attempts | % Scored |
|---|---|---|---|---|
| **1** | **Left** | 0 | 0 | NA |
| | **Right** | 23 | 53 | 43.40 |
| **2** | **Left** | 1 | 2 | 50.00 |
| | **Right** | 13 | 33 | 39.39 |
| **3** | **Left** | 0 | 2 | 0.00 |
| | **Right** | 9 | 20 | 45.00 |
| **4** | **Left** | 0 | 0 | NA |
| | **Right** | 11 | 31 | 35.48 |
| **5** | **Left** | 3 | 8 | 37.50 |
| | **Right** | 4 | 18 | 22.22 |
| **6** | **Left** | 11 | 23 | 47.83 |
| | **Right** | 7 | 13 | 53.85 |
| **5+6** | **Left** | 14 | 31 | 42.42 |
| | **Right** | 11 | 31 | 35.48 |

p_value of 5+6: 0.4373

Table 20: Scoring Percentage of Left and Right Handed Players From Each Position

| Position | Hand | Passes That Lead to Goals | Number of Passes | % Passes Let to Goal |
|---|---|---|---|---|
| **1** | **Left** | 0 | 0 | NA |
| | **Right** | 10 | 27 | 37.04 |
| **2** | **Left** | 0 | 0 | NA |
| | **Right** | 2 | 2 | 100.00 |
| **3** | **Left** | 0 | 2 | 0.00 |
| | **Right** | 4 | 6 | 66.67 |
| **4** | **Left** | 0 | 0 | NA |
| | **Right** | 24 | 58 | 41.38 |
| **5** | **Left** | 4 | 10 | 40.00 |
| | **Right** | 16 | 41 | 39.02 |
| **6** | **Left** | 15 | 37 | 40.54 |
| | **Right** | 6 | 19 | 31.58 |
| **5+6** | **Left** | 19 | 47 | 40.43 |
| | **Right** | 22 | 60 | 36.67 |

p_value of 5+6: 0.6914

Table 21: Scoring Percentage of Assisting Players That Are Left And Right Handed Players From Each Position

**3.5.6.1 Combining 5 and 6** We combined 5 and 6 in order to meet apply a z-test on the data, specifically we needed 10 goals for left and right handed in order to apply the z-test. While not as specific as doing 5 or 6 individually, the 5 and 6 are the only positions on the perimeter we expect to have left handed players. So combining the 5 and 6 allows us to see if having a left handed shooter/assisting player makes a difference.

### 3.6 Skip Shots

We compared the scoring percentage when players took skip shots versus normal shots, both overall and by position. The results can be seen in Tables 22 and 23 respectively. We also applied a two-sided z-test to the data we could, specifically overall, from 1, and from 4. Overall, the percentages were much better when the shot was skipped and the p-value is low enough for us to conclude that. For positions we can't conclude that for 1 and barley can for 4. We still need more data for us to feel comfortable making the distinction for positions. This implies skipping the shot is better; however, there could be other factors at play. Players may only skip when they either know the shot is going in or there is space in the water to skip, which may only happen if the defense is out of position already. Essentially, players may only skip when they already have a good shot.

| Type of Shot | Goals | Attempts | % Scored |
|---|---|---|---|
| **Non-Skip** | 49 | 162 | 30.25 |
| **Skip** | 33 | 60 | 55.00 |

p_value: 0.0007

Table 22

| Position | Shot Type | Goals | Attempts | % Converted |
|---|---|---|---|---|
| **1** | **Non-Skip** | 13 | 33 | 39.39 |
| | **Skip** | 10 | 20 | 50.00 |
| **2** | **Non-Skip** | 11 | 31 | 35.48 |
| | **Skip** | 3 | 4 | 75.00 |
| **3** | **Non-Skip** | 9 | 19 | 47.37 |
| | **Skip** | 0 | 3 | 0.00 |
| **4** | **Non-Skip** | 5 | 20 | 25.00 |
| | **Skip** | 6 | 11 | 54.55 |
| **5** | **Non-Skip** | 4 | 19 | 21.05 |
| | **Skip** | 3 | 7 | 42.86 |
| **6** | **Non-Skip** | 7 | 21 | 33.33 |
| | **Skip** | 11 | 15 | 73.33 |

p_value of 1: 0.4502
p_value of 4: 0.09997

Table 23

## 4 Future Work

In the future we hope to expand the work we are doing by gathering more data and exploring more ways to analyze the data.

### 4.1 Data

One of the main issues we have cited with our analysis has been the lack of data we have. In the future we hope to collect more data to help alleviate this issue. To help facilitate this collection, we want to make it easier to collect the data, whether through removing unnecessary data fields, improved cameras, or collaboration with more people. We also want to begin getting data on collegiate play, in order to expand the use of our analysis and compare how water polo changes across levels of skill. We also want to change how we collect data to help us better explore some of the analysis we have already started. We want to improve how we look at rotations. To do this we want get more exact with player location in the pool and find a metric(s) to help tie rotations of players with the movement of the ball. To expand how we look at left versus right handed players we want to determine the impact of having left/right handed players at certain positions, even if they aren't directly shooting the ball. We also want to add a field on the number of fakes players take before a shot and while passing to explore how faking with respect to different combinations of passes can influence the conversion rate.

### 4.2 Analysis

Analysis wise, there is lots of room for more in-depth analysis. That being said, there are aspects of the game we want to explore more than others. As mentioned above, we want to explore more into how different rotations allow the offense to generate better shots. We also want to dive deeper into some of the results that generated low p-values, skips and 3-3s, to see why they lead to such extreme differences in conversion rate and if they can better be used by polo teams. We are also curious on trying to figure out why the 5 position was consistently so bad. We also want to know is it inherent to the position or are there things a team can do to make it better.

## 5 Conclusion

After collecting data from the 2018 European's Champion League Tournament we have analyzed each 6-on-5 from a variety of different perspectives. While the data is limiting in many ways and more data is needed to be more definitive with our conclusions, we were able to come up with some takeaways that teams could use to improve their 6-on-5. The 5 position is clearly the weakest in the pool and should be heavily exploited by the defense. Only certain positions should pass to either the 2 or 3 post. If not going for the quick, teams should attempt to pass the ball sufficient amount times to open up better shots. Rotations by teams should be to the right. Giving up ejections in the back-court is probably worse for defense than giving them up in the front-court. Teams should try to experiment with ending their 6-on-5 in a 3-3 and using that to generate good shots. And if giving the opportunity, try to skip the ball.

## 6 Acknowledgments

We would like thank Bret Lathrope for introducing us the resources used and suggesting different ways to analyze the data. We also want to thank John Benedick, Tom Cronan, and Martin Desmarais for their input along the way. Additionally, we want to thank Kyle Archer for providing edits of this report.

## 7 Appendix

Information on data fields:
Num Passes - int
Passing sequence(i.e., list of positions the ball goes to) - list of int
Shooter Number - int [Note: Not recorded for half of games]
Shooter Position - int
Shooter Hand - 1 for righty, 0 else
Rebound - 1 if 6-on-5 is from rebound, else 0
Goal - 1 if resulted in goal, else 0
Block - can be possible strings 'none', 'goalie', 'field_player'
Skip - 1 if skip shot, 0 else
Lob - 1 if lob shot, 0 else
Shot Position - possible strings 'high', 'low', 'middle', 'not_on_cage' (case where shot goes over cage), 'none' (there is no shot)
Shot Location - possible strings 'left', 'right' , 'center', 'not_on_cage' (case where shot wide of cage), 'none' (there is no shot) [Note: 'left' and 'right' are from shooter's perspective]
Assisting Player Number - int [Note: Not recorded for half of games]
Assisting player position - int
Assisting player hand - 1 for righty, 0 else
Rotation Type - possible strings '6_in', '1_in', '2_pop', '3_pop', '2_in', '3_in', '4_center', '5_center', '1_pocket', '6_pocket'. These strings are put in a list
Timeout - 1 if timeout was called before 6-on-5, 0 else
Team - string
Turnover - 1 if ball was turned-over, 0 else
Quarter - int
Start time - minute:seconds
End time -minute:second

Dropped ball - int of number times ball was accidentally dropped
Kick out position - 1 if ejection was in back-court, 0 if ejection was in front-court
3-3 - 1 if ended in 3-3, 0 else
5 meter - 1 if lead to 5 m, 0 else
Link to play - string of link of play

# 8 Works Cited


[1] Osti, E. (n.d.). Retrieved May 29, 2019, from https://www.youtube.com/user/h2opolo69/

[2] Kim, E., Novak, L., & Shen, K. (2019, April 16). Water-Polo-Analytics. Retrieved from https://github.com/kevinshen1101/Wa
Polo-Analytics

[3] Kim, E., Novak, L., & Shen, K. (2019, May 29). Final_analysis. Retrieved from https://github.com/kevinshen1101/Water-
Polo-Analytics/blob/master/final_report/final_analysis.ipynb

[4] Mayberry, J. (n.d.). Exclusions: Efficiency over Frequency [PDF]. USA Waterpolo. Retrieved from https://usawaterpolo.org/do
over-frequency.pdf

[5] Ordóñez, E. G., Pérez, M. D., & González, C. T. (2016). Performance Assessment in Water Polo Using Compositional Data Analysis. Journal of human kinetics, 54, 143–151. doi:10.1515/hukin-2016-0043 Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5187968/

[6] Resources. (n.d.). Retrieved from https://usawaterpolo.org/sports/2018/11/28/resources-analytics-html.aspx