Design of an Expert System Coach for Complex Team Sports

Brice A. Colcombe, Lindsay M. Horton, Muhammad Ommer, and Julia Teng
George Mason University, bcolcomb, lhorton4, mommer, jteng2@gmu.edu

Abstract— **George Mason University Men's soccer team has only participated in the NCAA Division I tournament once between 2013 and 2015. During this period, the Atlantic 10 conference champions had an average win percentage of 51.6% and average RPI score of 0.69. Analysis of historic A-10 conference data illustrates that in order to win the conference in a regular season, a team must have an average win percentage of 0.79. Furthermore, the average difference in coaching salaries of NCAA teams that consistently appear in the championship compared to Mason's head coach is $67,457. This project describes the design of an Expert System Coach that will support the human coach by processing data from the events in games and then provides tactical adjustments through formation changes. The system processes game data to develop Pass Probability Maps for both teams. The system then utilizes heuristics based on emerging patterns to make tactical adjustments at halftime in different formations that yield new Probability Maps. These new maps will run again in the simulation until a formation is found with the best Goal Differential for each game. An analysis of 5 Mason games from the 2015 season did not provide enough data to accurately suggest tactical adjustments. However, by using all 5 probability maps, results became evident. Teams should aim for 70% passing percentage to increase their probability of winning. Lastly, goals are random events and games are often decided by their randomness, leading to coaches being put under unjust pressures due to performance.**

Index Terms – Expert System coach, Pass Probability Maps, Rating Percentage Index, Soccer

CONTEXT ANALYSIS

I. Soccer

Soccer is a game played with two opposing teams. Each team consists of ten field players and one goal keeper. The objective of the game is to score more goals than the opposing teams. A regular game is 90 minute long, the team with the most goals wins. If there is a tie then will be two 10 minute halves and the first team to score a goal will win the match.

The National Collegiate Athletic Association (NCAA) is the governing body that regulates specific set of rules in collegiate levels of sports including soccer [5].

Traditionally, the way soccer has been played was heavily based on the coach's expertise and experience with the sport. They are responsible in making decisions during matches. However, soccer has slowly been making the transition in utilizing data analysis for making important decisions, which is the focus of the Expert Coaching

System described in this paper.

II. Atlantic-10 Conference Statistics

Analysis of 2015's A-10 data showed that GMU averaged 12.12 shots/game and 1.35 goals/game [10]. This data can then derive an average "Goals per Shot" for the teams via the following formula:

$$\text{Goals per Shot} = \frac{\text{Goals per Game}}{\text{Shot per Game}} \qquad (1)$$

Figure I indicates the average performance levels between the A-10 teams in 2015. GMU averaged 11.11% goals per shot. In contrast to schools such as Saint Louis University (SLU), who averaged a higher 13.74% despite taking less shots (10.71 shots/game). SLU performed closer to Dayton with respect to Goals per Shot (at 14.01%), which is the clear outlier in Figure I. This illustrates the current performance gap between GMU and higher ranked schools as GMU's shots are not equating to goals at a high rate, which results in a decreased athletic performance of achieving higher ranks in the conference.
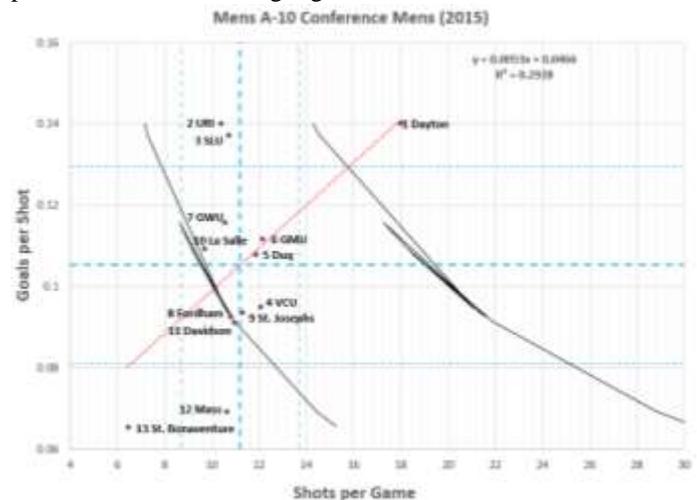

FIGURE I
2015 A-10 MEN'S CONFERENCE SEASON GOALS PER SHOT VS. SHOTS PER GAME

ROLE OF A COACH

Coaches are under constant scrutiny to perform and get results on the field to maintain their position and increase their salary. To be successful, a coach must be able to positively influence their players to commit to their coaching strategies and make any adjustments accordingly in order to follow the coaches' goal to defeat the opposing team. Coaches often rely on their expertise to spot any

cases where an adjustments must be made [2]. This is done in training and games. Coaches who only rely on this strategy of coaching are more traditional coaches and their current operations are modeled below in red in Figure II. With many new technologies out to aid coaches, lots of coaches are turning to a data driven approach. This type of coaching uses their expertise as well as data collected during training and games to make any appropriate changes. . This is modeled below in blue.
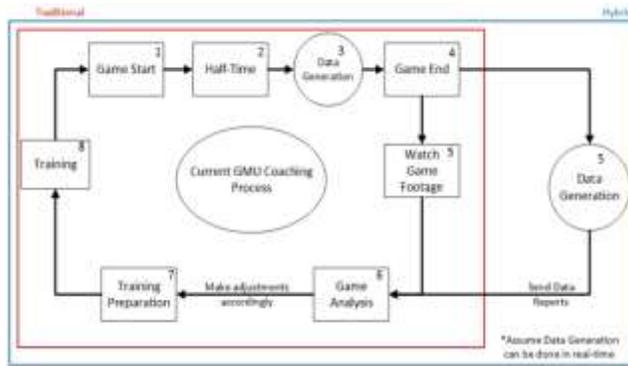

FIGURE II
HOW COACHING IS DONE

STRATEGIES

To follow with their set game plan, coaches have a variety of strategies which they hope to impose on their teams' to win games. These strategies often involve styles of formations that go along with them. Formations are how teams are set up on the field in order to have numerical advantages at certain places on the field. Some more popular formations include the 4-4-2 and the 4-3-3, both shown below in Figure III [4]. These formations both have different advantages and disadvantages at different areas of the field based on their numerical advantages. For example, a flat 4-4-2 has a lot of numbers at wide flank positions but may be outnumbered in the central midfield [4]. On the contrary, a 4-3-3 has a numerical advantage in central midfield but can be outnumbered at wide positions. With each formation, there are many variations that arise based on the style the coach implements. With a 4-4-2, coaches can play with a flat 4 midfield or a diamond midfield to have a numerical advantage in the central midfield. Small variations and changes are often made in game based on coaching expertise and numerical advantages in certain areas of the field [4].

FIGURE III
TWO TYPES OF FORMATION

STAKEHOLDERS

I. GMU Men's Soccer Team Coaches (Head/Assistant)

The role the coach's play is substantial as many of the decision-making criteria falls upon the coaching staff. The coaches are funded by GMU (specifically the athletic department) and train their soccer players to promote stronger performance in matches. Stronger performance leads to increase reputation for the respective school the team belongs to. As a result, the schools Division I team will make frequent appearances in the NCAA tournament, which will prospect to an increase in revenue for the represented school.

Ultimately, both head and assistant coaches want to develop the team to adapt and exploit the weakness of the opponent to win matches. Currently, GMU coaches rely on a hybrid strategy of traditional coaching formations as well as the use of data analytics reports provided by. At collegiate levels, this is done in order to receive an NCAA bid at the end of the season to participate in the NCAA tournament [3].

II. GMU Men's Soccer Team Players

This includes all the current members of GMU Men's Soccer Team. The players are recruited and then trained by the coaching staff during the regular playing season to win all matches. Their objective is to win matches consistently in order to increase their overall Rating Percentage Index (RPI) score for the respective season [3]**.** The players traditionally need to have a strong connection and trust with one another in order accomplish the objective of winning.

III. National Collegiate Athletic Association (NCAA)

The NCAA is a non-profit organization established in 1906 that promotes and regulates the athletic programs of colleges and universities [5]. This is done in order to maintain a set of standards that GMU Men's soccer team must abide by in order to promote fair gameplay. With these standard regulations come consequences derived from the NCAA. If college teams fail to meet standards, it can result in said consequences. As such, GMU is partnered with the NCAA committee to ensure that all teams have equal opportunities to succeed in their respective sports.

IV. Team Investors (George Mason University)

George Mason University works concurrently with the NCAA in order to establish universal rules and regulations that collegiate teams must abide by. These rules regulate what coaches can and cannot incorporate in their training programs. GMU also provides the athletic department with certain resources (i.e. scheduled training sessions, practice fields, equipment and salary) in the hopes that coaches will be able to develop a winning team, thus increasing the school's reputation in its respective athletic program.

GAP ANALYSIS

I. Atlantic 10 Conference History

George Mason University is a member of the 13-school Atlantic 10 Conference. Each year, the Atlantic 10 holds a conference tournament [1]. The winner of the conference gets an automatic bid into the NCAA tournament. Historically, Saint Louis consistently received a high average winning percentage of 62.5% in the conference. George Mason sits in 2nd place with VCU and Dayton in terms of win percentage at 50% over the past 3 years. However, the actual tournament champions during this time period had an average win percentage of 51.6% and collective RPI score of .69 [7].

II. NCAA Tournament

The goal of every collegiate team is to compete in the NCAA tournament. This tournament is a single elimination tournament featuring 48 teams [6]. 22 of the teams are the teams that have won their respective conference. The remaining 26 teams are determined through the team's rating percentage index (RPI). The RPI is calculated with a formula which consists of 3 aspects: (1) a team's winning percentage, (2) their opponent's winning percentage, and their opponent's opponents winning percentage [3]. Analysis of historical RPI data illustrates the importance of a strong strength of schedule when deciding which teams to play every year. The winning team takes the championship home.

III. Monetary Gap

Coaches who consistently win their respective conference championships are automatically accepted into the NCAA tournament. The coach that wins the NCAA tournament are substantially rewarded for their efforts through salary compensation. Analysis of head coaches salaries of GMU and other top schools reveal a monetary gap of about $67,457 per year [6, 8-14]. This number is derived both from coaching experience as well as respective performances of the schools.

IV. Performance Gap

A gap is present between George Mason University Soccer and perennial Atlantic 10 powerhouse Saint Louis University's Soccer Team [10]. To replicate Saint Louis's success over the past ten years, George Mason will need to win two out of the every five years Atlantic 10 championship. They will also need to receive an NCAA bid six out of every ten years. To accomplish this, they will need to have an average RPI score of at least .56 [8-14]. By closing the gap between the teams winning percentages, the head coach will be able to close the $67,000 salary gap as well.

PROBLEM AND NEED STATEMENTS

I. Problem Statement

GMU men's soccer team has a low average RPI score of 0.53 when compared to teams consistently playing in the NCAA tournament. As a result, the team has not been able to consistently receive an NCAA Tournament bid at a high rate, with only 2 bids in the last 10 years [10].

II. Need Statement

There is a need for an evaluation system that combines coaching expertise and game data that will help coaches understand the complexity of the game. This understanding will lead to a competitive advantage over opponents, which will increase the team's probability of receiving and NCAA Tournament bid to 3 every 5 years (60% of total bids).

CONCEPT OF OPERATIONS

The proposed concept of operations for the expert coaching system is described in Figure IV. This system will be used during half-time using data from the first half of the game.   The system will input the passing and shot data from the data generation and will be able to recognize patterns and process statistical data in order to output tactical adjustments that coaches are able to use in preparation for the second half as well as in preparation for any other upcoming games. This system will be able to 'see' the complexity that is often missed by the coaches, giving them a heavy advantage over their opposition.
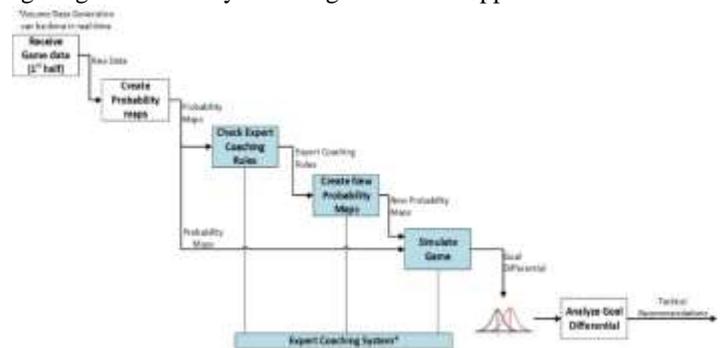

FIGURE IV
PROPOSED CONCEPT OF OPERATIONS

EXPERT COACHING SYSTEM

The Expert Coaching System utilizes George Mason's and their opponent's probability pass maps (PPMs) as inputs to the system of running the game simulation. PPM's determine how the ball moves in the 14 zones of our simulation. The system will run two parallel scenarios; (1) Continue with current formation for the second half of game simulation and (2) Utilizing statistics generated from first-half of the game and run them against the expert coaching rules. If the statistics do not meet the coaching rules passing percentages then a new strategy will be played (new set of probability maps) during the second half of game simulation. At the end of both simulated games, game statistics will be generated. The goal is to look at the score differential between two games to see if the new strategy increased the probability of winning the game.
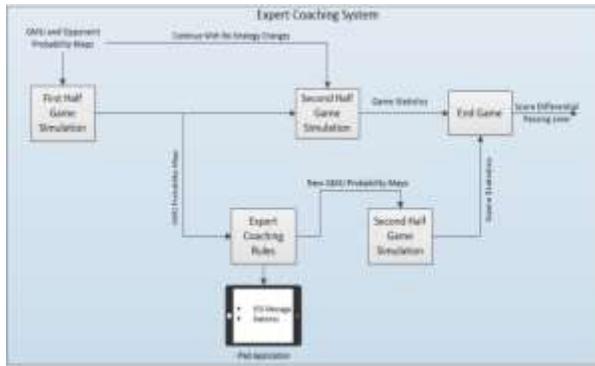
FIGURE V
PROCESS OF EXPERT COACHIING SYSTEM

SIMULATION

The expert coach system will simulate a ninety-minute soccer game based on previous game data. The code has been reverse-engineered from a previous senior design project that simulated referee proficiency based on soccer ball movements [15]. The simulation will run for the first half (forty-five minutes) before running against the expert coaching rules to see if a tactical adjustment are necessary. The tool is designed to aid the coaching staff in order to help them recognize complex patterns that occur during the game. Our simulation GUI is depicted in Figure VI and the algorithms are explained in the following sections.

FIGURE VI
SIMULATION INTERFACE (GUI)

I. Pass Probability Maps

The inputs into the simulation are a set of pass probability maps (PPM's) that combine to make one strategy. The probability maps are derived from five previously played game data. There are fourteen zones per map and there are a total of thirty probability maps that combine to create one strategy. Fifteen of the thirty maps are movement maps and the remaining fifteen are success probability maps.

The fifteen movement maps are comprised of the fourteen zone probability maps and one shot map. These fourteen zones in each map will combine to have a probability totaling to one. Below is a picture of zone grid and its respective probability map and how the zones are labeled. The fifteenth zone will be behind the opposing team's goal. Once a new zone has received the ball the movement probability map will change.

FIGURE VII
ZONE GRID AND RESPECTIVE PPM

The fifteen success probability maps are what determine if the pass or shot was successful. Each of the cells in the map will range from zero to one depending on the amount of success and failures from the input game data. For example if the movement probability map passed the ball from zone 0 to zone 1, the success map would find the probability if the pass from zone 0 to zone 1 would be successful or not.

The set of maps being used will change depending on what zone the ball is being shot or passed from and what strategy is being used. There are a total of 360 probability maps used for this simulation depending on the input into the simulation.

II. Ball Movement Algorithm

The ball movement will always go through the same algorithm in each zone. Once a ball has been successfully received in a zone, the player will have a choice to either pass the ball to a new zone or dribble the ball within the current zone. The algorithm runs a conditional. If the decision is to dribble the ball, then they must pass it after a maximum of 5 seconds. Once the ball has been passed, the algorithm runs a random number generator that that will act as a uniform distribution. The algorithm then runs through the correct movement probability map zones from zero to thirteen to determine what zone the ball is being passed to. Once the ball has been passed, another random number generator will be created. If the random number is larger than the success map probability selected for a given pass, then the pass will be a success. Otherwise a turnover will occur and the other team will have the ball. This is the **transitional** phase. If a shot action is selected instead of a pass, then the simulation transition will work in the same way as a pass. However, when a shot on goal occurs, the ball will always be overturned to the other team. The transition period is when a ball can be turned over to another team, which uses the opponents PPM's, or successfully passed to another zone. This results in a new

set of PPM's that the algorithm will now use since the zone of the ball or the team has been changed.

III. Expert Coaching Rules

The coaching rules were derived from interviews with the GMU's coaching staff. The rules follow the rules of each of the strategies. The field has been divided into sections; left wing zone (0-3), right wing zone (10-13), middle zone (4-9), which will have a cumulative value of one. There is also a defensive zone (0, 4, 5, and 10), midfield zone (1, 2, 6, 11, and 12) and offensive zone. (3, 8, 9, and 13) which will also have a cumulative value of one. Each rule is defined by a target total passing percentage of the total passes that occur in that zone. If the target percentages are not met, the ECS will send a message telling the user what strategy to play and why it decided that strategy to help give the team a competitive advantage. Table I depicts some coaching rules that the data runs though.

TABLE I
EXPERT COACHING RULE

Format ion		Right Wing Zone	Left Wing Zone	Middle Zone	Message	Tactical Change
4-4-2	Flat	≤30%	≤ 30%	≤ 40%	Get ball wide	4-4-2 Diamond
		>30%	> 30%	> 40%	Cont Strategy	None
	Diamond	≤25%	≤ 25%	≤ 50%	Find central players	4-3-3 Triangle
		>25%	> 25%	> 50%	Cont Strategy	None

SIMULATION VALIDATION

For simulation validation, we set all the movement maps to an equal value (1/15) and then changed the success probability maps. The success probability ranged from .1 to 1. When playing each other using the same success probability maps, the average possession for each team over all the trials was 50.14% for the home team and 49.86% possession for the away team thus validating that our simulation provided accurate results.

RESULTS

Through simulating games using the data which was available, it became evident that there was not enough quantitative and accurate data to use the coaching rules which make tactical adjustments. However, through summing up all the probability maps into one map, enough data was available to provide more accurate results. These summed maps were used to simulate GMU vs. GMU games which resulted in positive findings. The first finding is evident in Figure VIII which shows the goal differential during these games. While it is expected to be zero for every game, that is not the case. This suggests that a lot of

soccer games are decided by randomness. When two equally competitive teams play, the game can be decided by luck. With very low scoring games, goals in soccer are rare events which are hard to predict. Because of this, coaches are put under a lot of pressure to perform when in reality this simulation proves that their coaching does not impact the game as much as randomness.
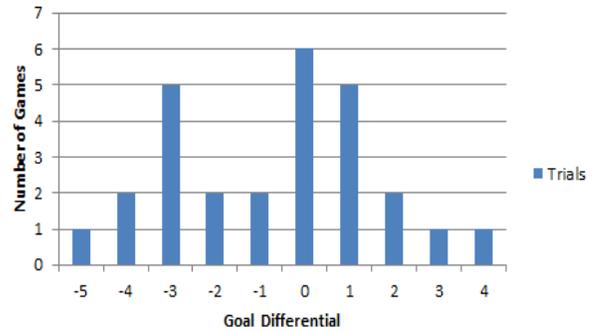

FIGURE VIII
GMU VS. GMU GOAL DIFFERENTIAL

When comparing shots and goals per game, a trend line can be fit to express the relationship. In Figure VIII, we can see the trend of the GMU's (whose success rates change from .1-1) compared to the away team (whose success rates stay constant at .5). It can be seen that both teams have roughly the same about of shots vs goals except for a few outliers in GMU's team. These outliers occurred when because they were able to take more shots, giving GMU more opportunities to score. Therefore when running this data, the threshold that proves most important occurs between exactly 60 and 70% success rates as seen in Figure IX. This is the tipping point that must be met in order to defeat an average team. As the level of competition increases (the success rate of the PPM increased) that threshold will also raise accordingly.
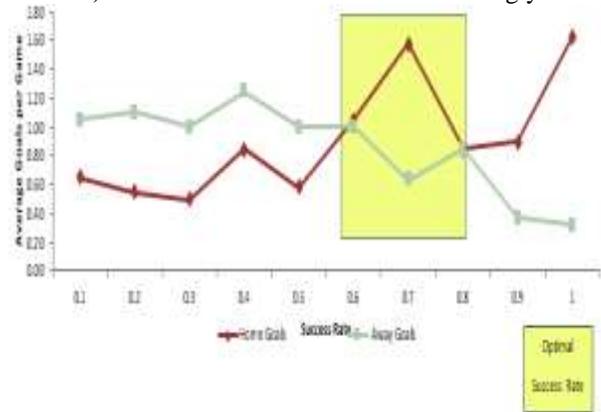

FIGURE IX
AVERAGE GOALS PER GAME SENSITIVITY ANALYSIS

By changing the passing probabilities, Figure IX also sheds light onto the result of increasing passing percentages. More possession leads to more shots, more shots lead to more goals. More possession also leads to a decrease in the opposing team's shots and goals. Because of this, a good offense is the best defense because it limits the opposing team's opportunities on goal.

CONCLUSION

There are many flaws in data collection that need to be fixed in order to run an accurate simulation. With these changes, our coaching tool can be used to positively affect the outcome of the game. Coaches should aim to have their team passing percentages at 70% to increase their probability of winning. Coaches and universities must understand that many games are won by luck and therefore coaches should be put under less pressure to win games.

RECOMMENDATIONS

With the results we obtained from our trails we can conclude our simulation functions as one would expect a soccer game to run. However, there is not enough data from previous games to create completed probability. In order to create a more accurate result, more precise data as well as additional game data must become available for use. The data we used did not give enough information for both teams where the passing in the field was occurring. At the times where there was the data showing where passes occurred, it was not consistent with some of the other data in the reports. It would also be useful if goal kicks, throw-ins and fouls could be recorded to help create a more realistic simulation. This will provide more insight into patterns and trends that occur throughout the span of a collegiate season. In addition, more layers can be added onto the simulation. While our simulation specified only the areas of the field of ball movements, each player can be implemented as another layer to add further complexity and accuracy to the simulation. This could add player fitness as well as certain player behaviors that are not captured in a probabilistic simulation. These advances will help to create complete movement and success PPM's which will help our team can successfully simulate games in order to find optimal strategies and optimal statistics that can give insight into changes coaches can readily make order to create a successful team on the field.

BUSINESS CASE RESULTS

Our startup company, Soccer Lytic's, will provide the ECS for NCAA Division I soccer coaches. The system will assist coaches in recognizing emergent patterns, thus increasing their respective competitive advantage in the game. Each year, there are 206 men's and 310 women's soccer team fighting for one championship. And on average, coaches whose teams consistently appear in the coveted tournament have a salary bonus of $67,457.30. Soccer Lytic's expert coaching system will assist coaches in solving this salary deficit. For our 1st year, our projected goal is a 1% penetration rate by obtaining 5 Division 1 men's or women's teams. The 2nd year, our goal is 50 teams (10% penetration rate). And by the 5th year, we hope to have 180 total teams at a 35% penetration rate of the available market. Soccer Lytic's will provide team-subscription packages, which cover all games throughout the season and access to online interactive dashboard. Each subscription will be $20,000 per year. Our starting cost for

Soccer Lytic's is $250,000 and recurring costs over 5 years will be $918,730 per year. With the cost and estimated unit sales each year, our company will be able to break even in about 3 years (depicted in Figure 10). In 5 years, we will have a cumulative revenue of $8,700,000.00. Soccer Lytic's will be able to earn $3,856,350.00 during this period.
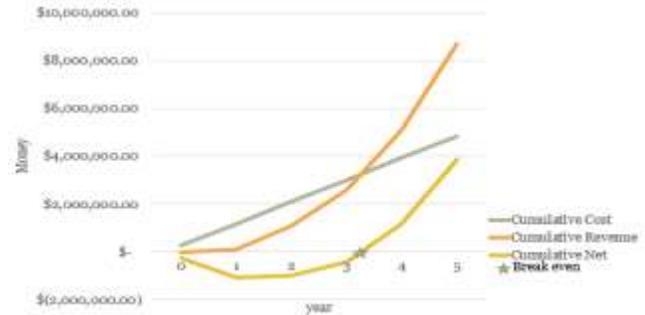

FIGURE X
CUMULATIVE COST VS PRODUCTION YEARS

References

[1] Atlantic10.com, Atlantic 10 Conference Official Athletic Site', 2015. Accessed: 07- Dec- 2015

[2] J. Perez, 'Concepts and Coaching Guidelines', U.S Soccer Curriculum, 2015. Accessed: 07- Dec2015

[3] Sites.google.com, 'RPI: Formula - RPI for Division I Women's Soccer', 2015. Accessed: 07- Dec- 2015

[4] Soccer-training-guide.com, 'The 4-3-3 is a Highly Popular Formation', 2015. Accessed: 08- Dec- 2015

[5] NCAA.com, 'NCAA.com "The Official Website of NCAA Championships', 2015. Accessed: 07- Dec- 2015

[6] Virginiasports.com, 'VirginiaSports.com - University of Virginia Official Athletics Website - UVA Cavaliers Men's Soccer', 2015. Accessed: 07- Dec- 2015

[7] '2009 Atlantic Coast Conference', 2015. [Online]. Accessed: 07- Dec- 2015

[8] Uclabruins.com, 'Men's Soccer - Schedule - UCLA Bruins Official Athletic Site | UCLABruins.com', 2015. [Online]. Accessed: 07- Dec- 2015

[9] Gomason.com, 'Men's Soccer - News - GoMason.com – The Official Athletic Site Of George Mason University', 2015. [Online]. Accessed: 07- Dec- 2015

[10] Gomason.com, 'Men's Soccer - News - GoMason.com – The Official Athletic Site Of George Mason University', 2015. [Online]. Accessed: 07- Dec- 2015

[11] Umassathletics.com, 'University of Massachusetts Official Athletic Site - Men's Soccer', 2015. [Online]. Accessed: 07- Dec- 2015

[12] Vcuathletics.com, 'VCU', 2015. [Online]. Accessed: 07- Dec- 2015

[13] Iuhoosiers.com, 'Official Athletics Site of the Indiana Hoosiers', 2015. [Online]. Accessed: 07 Dec- 2015

[14] Masslive.com, 'Massachusetts state employee salary database 2015', 2015. [Online]. Accessed: 07- Dec 2015

[15] N. Jones, A. Cann, S. Almashhadi and H. Popal, 'Assessment of Soccer Referee Proficiency in Time-Sensitive Decision-Making', George Mason University, 2012. [Online]. Accessed: 06- Dec- 2015