What makes a Successful Referee?

Determining and Evaluating the Characteristics of Successful Major League Soccer Referees

John Westbrooks

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Professional Doctorate

August 2022

Abstract

The value of professional soccer teams has been increasing year over year as the value of the players continues to increase. The increasing value of the teams puts more weight on each match. Soccer referees' decisions can have a major impact on the outcome of the match, influencing who wins and who loses. There has been little work examining the intricacies of what makes a soccer referee successful. The aims of this project were 1. Conduct study 1 to determine what the definition of a successful referee performance is and what are the characteristics of a successful referee from a broad range of stakeholders in Major League Soccer (MLS). 2. Use the results of study 1 in collaboration with PRO2 coaches to develop a way to assess the referees: a) to differentiate between different levels of match official and b) to determine areas of strength and weakness for each match official. 3. Develop a battery of off-field methods to assess determinants of referee performance, based on items deemed important during the Delphi study which: a) have a strong correlation with the in-game assessment and b) differentiate between distinct levels of official.

The first study used the Delphi methodology to ask MLS General Managers, MLS Coaches, MLS League Officers, PRO Referees and Assistant Referees, PRO Staff and PRO Assessors two questions: 1. Their definition of a successful referee performance. 2. Their opinion on the characteristics of a successful referee. The result of the first question was a 7-point definition of a successful referee performance: No one talks about their impact on the game/they do not negatively impact the game, they make very few errors/all major decisions are correct, they control the game, they correctly enforce the laws of the game, they protect the health and safety of the players, they communicate with the players and both teams feel like they received equal treatment. The second question yielded 26 characteristics of a successful referee: physically capable, correct, positioning, correct movement, proper use of body language, knowledge of the laws of the game/ability to apply the laws of the game, football understanding, strong communicator, ability to manage players/coaches, ability to anticipate play, teamwork, makes the correct decisions, experience, composed under pressure, focus, mentally tough, confident, adaptable, courageous, ability to lead, strong personality, approachable, committed, humble, mentally prepared, trustworthy, and student mentality. There were ten characteristics that overlapped with previous work examining successful referees.

The second study developed a match assessment tool for referees in collaboration with the coaches for PRO2. The coaches used the match assessment tool on 305 matches for the USL-C. The tool was able to differentiate between tiers of referee within PRO2 on 8 of the 18 items including total score. There was no difference in scores between coaches who did and did not coach the referee they were evaluating. This is the first project to evaluate the effectiveness of an in-game assessment tool and use it to differentiate between levels of official.

The third study attempted to differentiate between levels of official for the psychological and professional characteristics in study 1 as well as knowledge of laws of the game, communication and teamwork, and endurance ability. There was no difference between levels of official for the psychological and professional characteristics. There was no difference between level of official for laws of the game. There was a borderline difference between Tier B and Tier D for communication and teamwork. There was a difference between level of official for endurance ability measured by a 4-minute submaximal treadmill run, with MLS officials having a lower heart rate after 4min ($p = 0.001$), heart rate after 1 minute standing recovery ($p = 0.001$) and recovery heart rate ($p = 0.001$).. There were no off-field assessments that were able to reliably differentiate between level of official without another potential explanation.

This project was able to develop a definition of a successful referee performance and determine the characteristics of a successful referee. It was able to then use those characteristics to develop an on-field assessment that could differentiate between level of official. The development of an off-field assessment tool could not differentiate between level of official and did not possess a strong correlation with the on-field assessment tool. The results of this project can potentially be used to decide who should be promoted to a higher tier within PRO2, to drive coaching points for individual referees and to evaluate the effectiveness of current education methods.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application

for another degree or qualification of this or any other university or other institute of learning

Acknowledgments

<p style="text-align: center;">Table of Contents</p>

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 Professional Introduction

Understanding what it means for someone to be successful can be difficult. In many areas there are potentially conflicting definitions of success. In professional sports success is usually measured around an incidence of winning. Teams are successful if they win, and athletes are successful if they positively impact their team's chances of winning. It is far more challenging, however, to measure success when the athlete being evaluated never wins. How fast do you need to be if you are never trying to beat someone else to a spot or a ball? How fit do you need to be if you aren't trying to outlast someone else? These are questions that come to mind when training professional soccer referees. According to Helson and Bultynck (2004) a referee makes approximately 200 decisions during a match; do they have to get all these decisions correct to be successful? Could they do that and still be unsuccessful? It can be difficult to know as a coach if you are making the right decisions, if you don't know what goal you are trying to achieve and whether the person you are training is improving.

I started working with the Professional Referee Organization (PRO) in 2016 as an unpaid intern. This was my first experience with the idea of training and evaluating referees. Prior to this I thought they just showed up and blew the whistle. I did not understand the level of work that they put in prior to arriving to the field for a match. At PRO I eventually worked my up to a full-time role as a sport scientist primarily overseeing the implementation of our sport science program within the PRO2 group. PRO2 has 2 primary responsibilities: 1. Train and develop match officials for promotion to Major League Soccer (MLS). 2. Train and provide referees for the National Women's Soccer League (NWSL), United Soccer League Championship (USL-C) and League 1 (USL-L1) and MLS Next Pro. PRO2 is responsible for training approximately 80 match officials while assigning and evaluating close to 300 over all leagues. The officials within PRO2 are divided into 4 primary tiers, A, B, C and D with a 5th tier being created for prospects. The potential readiness for MLS matches is A to P in descending order of readiness.

I struggled to understand what my objectives were when preparing these match officials. I wasn't sure If what I was doing was helping them on the field or hindering them. I struggled to understand how we knew which officials deserved to be promoted or who needed to be demoted. I struggled to understand if our educational methodology was effectively improving

our officials. I wanted a way to understand what it means for a referee to be successful as well as a way to identify and track each referee's strengths and weaknesses.

I first developed the idea for this project prior to starting the professional doctorate program at Liverpool John Moores University. I conceived idea of creating a multifaced way of evaluating referees after reading Astroball by Ben Reiter in 2018. I felt the player profiles that they discussed could be a very effective way for us to understand what the talent levels of each official are, as well as individualize their coaching to target their areas of weakness. I was introduced to Dr. Barry Drust by my supervisor at the time, Matt Hawkey. I was intrigued by being able to work and complete the program at the same time. I especially liked the idea of being able to use the research structure to solve a program within my work. The structure of the Professional Doctorate made a lot of sense to me. It allowed me to keep developing my career as a practitioner while also developing another set of skills as a researcher. It also allowed me the freedom to ask and answer questions which excited me and had relevance to my career.

1.2 Research Introduction

Research in the field of refereeing has primarily been focused on understanding physical match performance, physical characteristics, and decision making during a match. There have only been four studies that have attempted to understand what makes a referee successful. Using qualitative methods Slack et al (2013) interviewed 15 English Premier League referees and found 8 higher order themes containing 24 lower order themes. The higher order themes were 1. Mental toughness attributes, 2. Support networks, 3. Effective game management qualities, 4. Multifaced pre-match preparation, 5. Performance-level enhancement, 6. Opportunities to thrive, 7. Personal characteristics, and 8. Superior physical components. Published in 2013, this was the first investigation into what factors underpin a successful referee. Pina et al (2019) interviewed 24 individuals with experience in and around soccer including referees, assistant referees, referee coaches, referee committee members, practicing and academic sport psychologists, observers, journalists, and physical and technical coaches. The 3 major categories they developed were individual preparation, game preparation and game management. Loghmani et al (2021) interviewed 12 Iranian FIFA referees and assistant referees. They found that officials who reached the elite level had more

experience playing soccer before starting their refereeing career. They found that support of family, specifically the father, and peers played a major role in these officials starting their careers. Another unique finding in the early period of the career is the importance of goal setting and individual practice for these officials successfully progressing up to the elite level. Mendes et al (2021) looked at referee data for a ten-year span between 2009/10 through 2018/19. They found that length of time on the national board, age, previous history of playing, and proximity to a dense population of other national board officials were predictors of an official having a higher ranking. These studies provide some guidance on what the characteristics of successful referees are in four different leagues, but further work is needed to better understand what factors underpin successful referee performance, especially in Major League Soccer (MLS), in order to facilitate optimal referee development and performance

Considering the relevant previous work and looking to expand on gaps in that research, the aims of this project were 1. Conduct study 1 to determine what the definition of a successful referee performance is and what are the characteristics of a successful referee from a broad range of stakeholders in Major League Soccer (MLS). 2. Use the results of study 1 in collaboration with PRO2 coaches to develop a way to assess the referees: a) to differentiate between different levels of match official and b) to determine areas of strength and weakness for each match official. 3. Develop a battery of off-field methods to assess determinants of referee performance, based on items deemed important during the Delphi study which: a) have a strong correlation with the in-game assessment and b) differentiate between distinct levels of official.

Chapter 2

Literature Review

2.1. Introduction

A soccer referee is an athlete with a unique set of demands expected of him or her during a match, both physically and mentally. The current duties of a referee according to the International Football Association Board (IFAB) are to "1. Enforce the laws of the game. 2. Control the match in cooperation with the other match officials. 3. Act as timekeeper, keeps a record of the match and provide the appropriate authorities with a match report including information on disciplinary action and any other incidents that occurred before, during or after the match. 4. Supervise and/or indicate the restart of play". The performance of these duties leads a referee to make approximately 200 decisions during a match (Helsen & Bultynck, 2004). Ultimately any one of these decisions can have an influence on the outcome of the match. The market value of players in the top 4 European leagues has increased by over 200% between 2005 and 2018(R Bloggers, 2020). This means the value of each game and therefore the potential financial impact of each decision increases. Thus, decision-making is one of the most important aspects of their role as referee. It is therefore important to understand referee decision making.

There are observable decisions (where the referee blows the whistle, or a clear event occurs) and unobservable decisions where the referee does not interfere with play. According to Helsen & Bultynck (2004) elite referees make on average 137 observable decisions and 200 total decisions in a match. This leads to an average of 3-4 decisions every 60 seconds of effective playing time. Castillo et al (2019) found that elite referees made an average of 225.3 decisions per match over 16 matches of an international tournament. The error rate of the decisions made by field referees has varied from 5% (Castillo et al., 2019) to 6.9% (Hossner et al., 2019) 14% (Mallo et al., 2012) to 36% (Mascarenhas et al., 2009). One caveat to the studies by Mascarenhas et al (2009) and Mallo et al (2012) is the smaller number of games and officials observed; 7 games/5 officials and 15 games/10 officials, respectively. In contrast, Hossner et al (2019) (64 games,25 officials) and Castillo et al (2019) (63 games,20 officials) observed a larger number of games and more referees. It is possible that the decision accuracy in Mascarenhas et al (2009) and Mallo et al (2012) would have been comparable to Hossner et al (2019) and Castillo et al (2019) with a larger sample of games. These decisions can be influenced by different factors such as crowd noise (Goumas, 2014; Nevill et al., 2017; Picazo-Tadeo et al., 2017) player/coach interactions (Lex et al., 2015) communication from

teammates (Boyer et al., 2015), distance from play, angle of view, their level of fatigue (Weston et al., 2012), time of the match, score (Lago-Peñas & Gómez-López, 2016), and the type of decision. Referees make many decisions during a match with several potential factors which could influence the accuracy of their decision. This places considerable mental demands on the referee. In addition to these mental demands the referee also needs to be able to handle the physical side of refereeing.

To make decisions a referee has to spontaneously move around the pitch at varying speeds and in varying trajectories to obtain optimal positions to make decisions; consequently, there are considerable physical demands placed on a referee. A referee will cover approximately 11km during a match while typically being 10 to 15 years older than the players (Weston et al., 2012). In addition to covering greater than 10 km during a match, referees will perform almost 900m of high-speed running (>19.8 km/hour) and an average of thirty sprints (>19.8 km/hour) reaching an average top speed of 31.8km/hr (Weston 2011). The physical demands of the match presented by these data result in a large physical challenge to the referee. If the referee is not properly prepared for this challenge the result could be an injury. There have been several studies looking at the injury rates per 1000 hours for referees during training and matches. The injury rates found per 1000 hours of training were 4.6 (Kordi et al., 2013), 0.9 (Bizzini et al., 2011) and, 2.16 (Paes et al., 2011). The injury rates found for matches per 1000 hours were 20.8 (Bizzini, Junge, Bahr, Helsen, et al., 2009), 34.7 (Bizzini, Junge, Bahr, & Dvorak, 2009), 19.6 (Kordi et al., 2013), 2.06 (Bizzini et al., 2011), and 2.16 (Paes et al., 2011). The most common injuries were to the lower leg and consisted of strains or sprains (Bizzini et al., 2011; Bizzini, Junge, Bahr, & Dvorak, 2009; Bizzini, Junge, Bahr, Helsen, et al., 2009; Kordi et al., 2013; Paes et al., 2011). The higher injury rates found during matches could be a result of the physical demand the match places on the referee.

Weston et al (2010) shows the importance of consistent physical training by laying out the physical training of an elite international referee over an 8-year period including a World Cup and UEFA Champions League final. This can be used as a guide for other match officials, as this official missed only two matches due to injury during that 8-year period. The variety of intensities of movement of the referee during a match creates the need for training which targets multiple levels of intensity from slower aerobic work to higher intensity aerobic work

to bouts of repeated sprints and maximal-effort sprinting. The physical and cognitive demands of the match together play a role in influencing the decision making of a soccer referee.

Given the importance of decision making for referees, for the purpose of this literature review the rest of the review will be split into 2 parts. The first portion of the literature review will be structured around the decision-making process model developed by Samuel et al (2020) to discuss the how the various aspects of referee performance relate to successful on-field decision-making. The second portion of this literature review will examine off-field aspects that contribute to a successful referee.

The decision-making process model provides a useful framework to explore referee decision-making and the factors which effect it. Samuel et al (2020) surveyed 13 referees and 7 assistant referees for their opinion on the model. The referees felt that the model accurately described their decision-making process, included the critical components and was clear and concise. The model from Samuel et al (2020) contains five levels of decision-making when a referee is evaluating whether to intervene on the field. The first decision is determining where on the field they need to move to have the best vantage point. This will relate primarily to the official's physical capacity, soccer understanding and the importance of position and angle of view. The second decision is filtering the relevant cues from the irrelevant cues to help them anticipate what will occur next. This will relate to physical visual abilities, visual processes, and experience level. The third decision is then to identify the event that occurs and prepare their possible actions to execute next. This will relate to prior training to recognize misconduct situations. The fourth decision is matching what they observed to the criteria within the laws of the game and then filtering it through the considerations for game management and coming to an accurate decision. This will relate to knowledge of the laws of the game, models for game management, and the influence of bias. The fifth decision is absorbing the feedback they receive after they make their decision and determine if they need to keep their decision or consult their teammates to change their decision. This will relate to communication with teammates, video assistant referee (VAR), interaction with players/coaches, and the dynamic aspects of refereeing as an emergent task. The following section will further elaborate on each stage of the decision-making process according to the model of Samuel et al (2020) and will review relevant literature.

2.2. Referee Decision Making Process

2.2.1. Determining Optimal Pitch Position

Many referees tend to follow a diagonal pattern of movement on the field. This line goes from the corner of penalty area opposite their assistant referee on one end of the pitch to the opposite corner on the other side of the pitch. The first decision that a referee needs to make is to determine if they need to move off this diagonal to have the best vantage point for the potential incident. Distance from play has been examined by Hossner et al (2019) during the 2014 FIFA World Cup. They found that when a referee was 10-15m from an event they were 2.58 times more likely to make a whistle error (incorrectly whistling for an infraction when none occurred) while inside of 5m they were 5.51 times more likely to make a non-whistle error (not whistling when an infraction occurs). It is possible that the officials felt more comfortable whistling when they were closer to the average distance from event, 13.7m, which could contribute to a high error rate at this distance. They found these officials were more likely overall to make a non-whistle error vs. a whistle error; 23.7% and 4% respectively. This means that when an official decides to blow the whistle for an infraction it is very likely that they are correct in doing so. Contrary to De Oliveira et al (2011) who found an ideal viewing distance of 20-25m, and Mallo et al (2012) who found an ideal viewing distance of 11-15m, Hossner et al (2019) did not find an ideal viewing distance. Castillo et al (2019) divided the field into ten sections and compared the accuracy of the decisions made in each of these sections. While location on the pitch did not affect the accuracy of decision, they found the decision type with the highest error rate was simulation, where a player acts as if they are fouled when there is no contact with the opponent. The research is unclear if there is an ideal distance from play for an official to make the correct decision. However, the referee still needs to have the physical capacity to consistently get to where they need to go on the field throughout the duration of the match.

There have been numerous descriptive studies focusing on physical performance during matches, primarily with top level male referees. The physical workload of a soccer referee is related to the physical output of the players (Weston et al., 2006; Weston et al., 2008; Weston et al., 2011). The referee will typically cover between 9 and 13km during a match. (D'Otavio & Castagna, 2001;Bangsbo et al., 1991; Weston et al., 2012; Weston et al., 2007; Malo et al.,

2009). While covering this distance the referee will typically work at 85% to 90% of their maximum heart rate and 70% to 80% of their VO_{2max} (Catterall et al., 1993; Castagna et al., 2002; Johnston & McNaughton, 1994;Krustrup & Bangsbo, 2001). The number of high-speed actions performed during a match is the best differentiator between levels of referee fitness (Mallo et al., 2009). These high-speed actions are made up of high intensity work and sprinting actions, which along with total distance covered are directly associated with the movements of the players during the match (Castagna et al., 2004;Weston et al., 2007;Krustrup et al., 2009;Mallo et al., 2009;Barbero-Álvarez et al., 2012;Costa et al., 2013). The referee mirrors the work of the players while in some cases being almost double the age of the players. While age can affect the physical output ability of the referee as shown by Weston et al (2010) their experience allows them to make up for their physical decline. The right type of physical preparation is important to make sure referees can withstand the demands of the match.

 In order to understand if an official is prepared for the demands of the match, it is important to be able to perform a physical assessment prior to the start of the season. The number of studies investigating testing and training of referees has been increasing to try to better align the physical testing with match demands and better prepare the officials for the match. Referees are not able to start a season without being match fit. Unlike players they cannot be subbed off part way through the match as they work to regain match fitness. They need to be physically ready to handle the match demands discussed above. There has been limited work on creating fitness tests that accurately represent the demands of a match. A review by Weston et al (2009) examined the data looking at the construct validity of the 150/50 interval test consisting of running 150meters in 30 seconds then walking 50meters in 30 seconds repeated for 10 laps around a track and 6x40meters with 90second recovery version of the FIFA Fitness test. Weston et al (2009) found that heart rate load, calculated by multiplying the duration in each zone (1-5) by a multiplier for each zone and then adding them together, during the 150/50 had a very large correlation with total distance covered during a match (r=.70) and a large correlation with high-speed running during a match (r=.57). The correlation between high-speed running and test heart rate with this test was close to or larger than the previous test used by FIFA, the 12min run, where the official must run as far as possible around a track in 12minutes, (.46, .54 and .51, respectively). There has not been any work evaluating

the newer form of the FIFA Fitness Test where the official must complete 75m in 15s followed by a 18s recovery period where they walk 25meters repeated 40 times.

Much of the research has examined the Yo-Yo Intermittent Recovery Test Level 1 test. It had a very large correlation with high-speed running during a match (r=.77). Casajús & Castagna (2007) found that the YYIR1 was able to discriminate between elite and sub-elite referees (p<.05). Weston et al (2009) found that the old 6x40m sprint with 90 seconds rest was highly related to both high-speed running during the match r=-.76 and sprint distance r=.77. Meckel et al (2020) compared first and second division referees on the newer 6x40m with 60 second rest but did not evaluate the relationship to match performance. Riser et al (2018) found no association between the 6x40m with 60 second rest and match performance for field referees and no association between the change of direction ability (CODA) and match performance for assistant referees. Riser et al (2018) found a weak association with 5x30m sprint and High-Speed Running (>19.8 km/hour) bouts during a match but no other metrics.

Bouzas-Rico et al (2021) reviewed 13 studies examining field-based fitness tests for referees. They found that there were no studies that evaluated the psychometric properties of a physical-cognitive test with decision-making. They found that only the Assistant Referee Intermittent Endurance Test (ARIET), similar to the Yo-Yo test with the addition of lateral shuffling and the 12-minute run test had good results for criterion validity and absolute reliability, but the 12-minute run test has poor content validity. They found the 50m sprint to be the best option for assessing maximal speed with strong absolute, convergent and construct validity scores. However, Mallo et al (2007) has shown that the 50m sprint is a poor predictor of referee performance during a match. There is still work to be done to find an adequate way to determine if an official is ready to meet the physical demands of the match.

One area that is missing in the literature is the effect of experience and soccer understanding in differentiating an official's ability to decide where to move on the pitch. On average, older officials cover less total distance and less high-speed distance as well as performing fewer sprints during a match than younger officials (Weston et al 2010). Therefore, it seems likely there is change in their movement strategy facilitated by an increased ability to understand where they need to be and to anticipate the actions of the players. More work is needed to

expand on this aspect. Once the official has decided where to move on the field and arrived there, they then need to decide what cues are important to focus on.

2.2.2 Filtering cues to anticipate upcoming events

The visual abilities (ability of the eye to see) of a referee and the visual strategies (how and where the eye is focused) they use are two aspects that can separate elite officials from sub-elite officials. Ghasemi et al (2011) found that referees who were most successful in adjudicating video clips from matches scored significantly better on visual tests examining the visual abilities of accommodation (ability to see things at different distances), saccadic movements (how quickly the eye can change from focusing on one object to another), recognition (how quickly an object can be recognized), visual memory (recall of an object seen) and peripheral vision (ability to utilize information from the peripheral visual system) than the less successful referees. Baptista et al (2021) found that visual ability of soccer referees, as measured by a test of visual perception skills, was associated with high on-field performance as ranked by the Portuguese football federation. Visual strategies accounted for 22% in the variance of on-field performance.

The visual strategies adopted by elite officials is another way that they can separate themselves from sub-elite officials. Spitz et al (2016) found that elite referees have a different visual search pattern while watching video clips compared to sub-elite referees, potentially explaining the difference in correct responses between elite and sub-elite referees during the video clips. The elite referees spent less time on non-relevant areas of the clip compared to the sub-elite officials. Moreira et al (2020) found that the duration of gaze fixation was significantly less in national level referees compared to state referees with no decrease in accuracy during a video-based test monitored with eye-tracking software. These findings suggest that elite referees can identify task-relevant cues more quickly and filter out irrelevant information. This could be the underlying reason why Catteeuw et al (2009), Gilis et al (2006) and MacMahon et al (2007) have shown a difference in decision-making accuracy between different levels of refereeing expertise. The experience of these elite referees may have taught them what cues are important to pay attention to and what cues to discard allowing them to quickly make an accurate decision. They can focus on the correct area based on external

stimuli to observe important information to get the decision correct. It is also possible that there is a baseline level of visual ability which is needed to become an elite official.

There may be an optimal combination of visual ability and visual strategy which leads to accurate decision-making. The visual ability and visual strategy developed over time have allowed the referee to decide which cues are important and next they need to identify the event(s) that has occurred.

2.2.3. Identifying the event that occurs and preparing their possible actions to execute next.

One area of focus in decision-making research is the use of videos to test and train soccer referees. Spitz et al (2017) found superior performance for elite officials during a video-based test compared to the performance of sub-elite officials. They found the elite referees were able to better anticipate foul situations and more accurately recall the exact location of the foul. They found that decision-making accuracy, anticipation and recall accuracy were positively associated with elite referee status. In addition to evaluating referees via video-based tests there are studies examining the effectiveness of utilizing video to train match officials. Nurcahya et al (2021) found video-based training effective for improving the performance of assistant referees compared to no training. This agrees with the work by Armenteros et al (2019) who found on-field training session performance improved after six interactive video sessions. Kittel et al (2019) performed a systematic review of the literature and found that video testing was an effective way to differentiate skill levels of soccer referees. They discuss the limitations of video testing mentioned in several of the studies i.e., lack of crowd noise, lack of physical exertion/fatigue and only one angle of view that is not the exact view of the official. They question how much the score on a video test corelates to on-field performance, but so far, no soccer referee studies have examined this.

When it comes to training or evaluating referee decisions, video speed can also play a role in accuracy. Spitz et al (2018) found that when a referee viewed a clip of a foul situation in slow motion, they penalized it more harshly than when they viewed it at normal speed. These results differ from those seen by Mather & Breivik (2020) who found red cards were judged more severely while yellow cards were judged less severely when clips were played at slower than normal speed. This has implications for both the evaluation and training of officials. One may be tempted to watch a clip in slow motion to try to better judge it, but certain stimuli

(i.e., force) may be lost when the speed is decreased. Elite officials can better recognize the offense that is being committed and anticipate what the decision is going to be allowing them to make a decision more accurately. Once the official has decided what event has occurred then they need to filter the observed event through the laws of the game vs. game management.

2.2.4. Matching observations to the laws of the game while considering game management to aid decision accuracy.

Scoppa (2021) and Mascarenhas et al (2005) described game management as referees only stopping play when incidents occur that demand intervention and otherwise letting the game flow. Work by Raab et al (2020) aimed to create a model to explain the differences between officials as to when they decide to utilize game management vs. applying the laws of the game. They propose that each referee has a threshold above which they go from simply applying the laws of the game to utilizing game management to keep control of the game. They believe that this threshold can be predicted using the Preference for Intuition and Deliberation Scale (Betsch, 2004). Using this threshold, one could then predict whether a given official would call a foul and give a caution or simply call a foul in a particular situation, and then how that situation would affect their likelihood to caution or not in future situations. This model could fit into the decision-making model by Samuel et al (2020) to describe the way the officials make the fourth decision in the five-decision model.

There have been a few studies that have examined the effect of this relationship between applying the laws of the game and game management. The concept of game management was elucidated by Unkelbach & Memmert (2008) who found that officials watching a replayed match in order (0-90 min) administered fewer yellow cards than the referees who watched the replayed match in a randomized order. They found that this could be due to the referees who watched the match in order decided to not punish a challenge early in the match compared to the referees watching in randomized order who punished offences in isolation. They also found that early in the match the officials watching in order set the scale for determining what would be a cautionable offense in the match. Game management can also explain why Plessner & Betsch (2001), and Schwarz (2011) found when a referee gives a penalty to one side, they are more likely to award a penalty to the other side. It is possible

they reduce their criteria for a penalty to help manage player emotions during the match. This can provide a feeling of consistency or fairness. Ultimately a referee switches between applying the laws of the game and game management for each decision based on what they observe relative to the laws of the game, e.g., some decisions are mandatory, and filter this through what decisions they have already made, what the context of the game is at that moment and then make their decision.

One factor which could affect the decision between applying the laws of the game and applying game management is the bias of the referee. That is, it appears that the referee favours one team over another in the decisions they make. There have been several papers that have shown that referees are not 100% objective in decision making (Dohmen & Sauermann, 2016). This bias has tended to be shown to benefit the home team. The effect has been an increase in number of fouls in favour of the home team, (Armatas & Pollard, 2014;Liu et al., 2019;Nevill et al., 2002), yellow cards given to the away team (Armatas & Pollard, 2014;Downward & Jones, 2007;Goumas, 2014;Liu et al., 2019;Sutter & Kocher, 2004;Unkelbach & Memmert, 2010), red cards given to the away team (Armatas & Pollard, 2014;Liu et al., 2019;Pettersson-Lidbom & Priks, 2010), and penalty kicks for the home team (Armatas & Pollard, 2014;Liu et al., 2019). Additionally, referees add less time if the home team is ahead by one goal, or more extra time if the home team is behind by one goal (Garicano et al., 2005;Sutter & Kocher, 2004).

A few papers published recently have compared matches with a full stadium to matches with an empty stadium as a result of the COVID-19 pandemic. The results from Scoppa (2021) found that differences between fouls, yellow cards, red cards went from favouring the home team to being neutral in matches without fans. McCarrick et al (2021) found that when they controlled for playing style of the team (e.g., the away teams were more aggressive/defensive) the occurrence of home team bias by the referee disappeared. Fischer & Haucap (2021) found a reduction in referee bias via an increase of 0.6 yellow cards and 1.1 fouls per match for the home team. However, they found this change to be too small to explain the entirety of the disappearance in home field advantage. This calls into question the impact of referee home team bias on the existence of home field advantage when fans are present in the stadium. Sors et al (2020) found that referee bias in favour of the home teams for fouls, yellow cards, red cards, penalty kicks and extra time disappeared when fans were not present. They found

the most significant effect for the number of yellow cards issued in favour of the home team. Sors et al (2020) posit that the social pressure of the crowd creates an unconscious bias in the referee for the home team, contributing to the home field advantage when crowds are present.

While a referee may feel that they are impartially switching between applying the laws of the game and applying game management, recent research has shown that the crowd can have an effect on how they switch between the two. The social pressure applied by the crowd can influence them to make slightly more decisions which favour the home team compared to the away team. However, when the crowd is removed then this social pressure does not exist, and the decisions become more balanced. Once the official has decided to apply the laws of the game or game management, they then receive feedback based on their decisions from teammates, players, coaches, and the fans.

2.2.5. Absorbing the feedback after a decision and determining if to keep their decision or consult their teammates to change it.

A referee's decision making can change within a match, as well as from one match to the next match based on the feedback they are receiving on and off the field. Russell et al (2020) examined the dynamic nature of referee decision-making consistency. They interviewed 8 Australian referees to attempt to understand their decision-making behaviour. They found that referee decision-making is an emergent behaviour that grows and develops throughout the duration of the match. They found the referee's view consistency as relative to within the game, not necessarily from game-to-game as the context and what is presented/needed within each individual game changes based on the players. It does not always mean responding to isolated incidents in the exact same manner as the incidents occur as part of a larger dynamic environment.

This work fits into the relationship between decision 4 and decision 5 of the Samuel et al (2020) model. It is the interaction point of a referee observing the response to their decision and seeing if it achieved the purpose they were looking to achieve. If they achieved their aim then they maintain their decision-making and if not, they may need to adjust their decision-making to achieve their aim. It also builds on the idea from Raab et al (2020) that there is a threshold for each referee from where they switch from applying the laws to managing the

game. That threshold is influenced by the context of the game, what the players need from the referee, where they had placed their previous line of intervention, and what purpose the referee is trying to achieve with a decision. These variables will change throughout the course of a match and with it the referee's decision-making will adjust. This has implications for the evaluation and the training of referees. The referee may also adjust their decision-making based on feedback from their teammates. One of these teammates is the recently introduced position of VAR.

The VAR was added to the referee team to help reduce errors in clear match-changing incidents, specifically goals, penalty decisions, red cards, and mistaken identity. They are a certified referee who watches the game with a video-monitor allowing them to review potential errors during the match and communicate with the on-field crew. Spitz et al (2021) examined the use of VAR during the 2016/17 and 2017/18 seasons in 13 countries over 2195 matches. There were 9094 situations checked where there was a clear correct outcome. Pre-VAR, the referees were correct 92.1% of the time, the use of VAR moved the accuracy up to 98.3%. The implementation of VAR has allowed for an improvement in the decision-making accuracy of the referee team. The referee is able to correct decision errors made on the field. It is yet to be evaluated how that affects the decision-making process of the referee when they are making their initial on-field decision. It is possible that knowing they have the VAR to correct their mistakes could lead them to make decisions differently than they did before VAR was introduced.

2.2.6 Other reasons for decision errors at any stage of the process.

It is possible that the referee is purposefully making incorrect decisions for reasons that lie outside of those outlined above. Visschers et al (2020) found that out of 595 Belgian referees surveyed, 44% believed that at least 1 out of 10 matches are fixed and 25% believed that they have witnessed match-fixing occurring. All of these officials surveyed were at the grassroots level. While the perception of match-fixing may be higher than the amount that occurs, it is still a point of major concern for the Belgian federation. One of the primary reasons for concern was the Portuguese scandal in 2004 involving officials and the referee administration (Moriconi, 2020). Moriconi & De Cima (2021) provides a sociological basis for why referees participate in match-fixing. This can be because the structure in place punishes those who

speak out by shortening their career. It could be that there is no place to submit complaints. Or it could be that the informal and formal rule structure prevent those who participate in match-fixing from being dealt with effectively. The history of match-fixing in soccer is well documented (Moriconi & Teixeira-Diniz, 2016). Referees have been bribed with gifts, money, sex and it has transitioned over time to positive evaluations and appointments (Moriconi & Teixeira-Diniz, 2016). The authors discuss the need to implement a system that targets the entity perpetrating the corruption, not just those carrying out the individual cases of match-fixing. Hill (2010) discusses the three pieces needed for match-fixing to occur on a large scale: large illegal gambling networks, relative depravation of players, coaches, or officials, and finally the expectation that cheating occurs.

2.2.7 Effects of Travel and Climate

Referees in MLS have an additional stress that referees in smaller countries do not have. The teams and referees in MLS are spread out between two countries, 4 time zones and ~4800km. In addition, the climate or weather conditions of a match can vary greatly depending on the time of the year. A referee may be officiating on one coast one week and another coast the next week. Early in the season they may be officiating in ambient temperatures below 0°C and by mid-season be officiating in temperatures above 35°C. Furthermore, they may be officiating at sea level one week and 1600m the next. Such adverse environmental conditions can negatively affect physical and perceptual performances of athletes.

Almost all the matches officiated entail air travel to reach the destination. There has been very minimal or no work examining the effect of time zone changes and distance travelled on referee performance. However, a review by van Rensburg et al (2021) highlights the increasing impairment of mental and physical performance of athletes the more time zones they cross to reach their competition location. The impact is different depending on the direction of travel with some work showing that east to west travel has a greater negative impact on performance (Fowler et al. 2017). These performance impairments are potentially caused by a mismatch in circadian rhythm with sleep/wake cycle required by the destination, dehydration caused by air travel, muscle stiffness and blood pooling from being seating in an airplane, and lack of sleep on long distance air travel. Despite the clear effects of single trip travel, one area that has not been considered is the accumulated stress of repeated travel.

MLS officials are away from their families for typically two nights. Air travel usually involves a 20-60 min drive to the airport, getting through security, sitting at the gate, the flight itself, getting out of the airport to either the shuttle or the rental car and finally checking into the destination hotel. This is performed 25-30 times during a season. While the officials may adapt to the lifestyle it still poses an additional stress on them prior to the match. These additional external variables add to the difficulty in evaluating referee decision making in the United States.

Examining the literature surrounding referee decision-making through the lens of Samuel et al (2020) displays how the act of refereeing is multi-factorial. A single decision encompasses five different decisions which can be influenced by the referee's physical capacity, hydration status, pre-match nutrition, reading of the game, visual ability, visual search process, ability to identify the event, experience, context of the match, threshold for management, threshold set for intervention, interaction with players, feedback from teammates and VAR, crowd noise, and previous decisions made in the match to that point. The combination of these factors makes it difficult to identify what makes a referee successful in their decision process. Specific work is required to understand what makes an official successful.

2.3 Part 2 Successful Referees

Referee decision-making and the contributing factors are very important components to a successful referee. However, it is also important to understand what else contributes to successful referees, and how to evaluate successful referees. Such an understanding will allow greater insight into what causes referees to decide to quit, how to support officials structurally and how multiple factors come together to create a successful referee.

2.3.1 Referee Recruitment and Retention

In the United States referees are recruited initially by the states to initially work youth matches as a grassroots referee. They are recruited through friends or local clinics sponsored by the state. The recruitment efforts depend in large part on the need within the state. The official progresses through their career by steadily working older age groups. Once the criteria have been met, they can become a regional referee and work matches outside of their state. They can then apply to be a national referee to work matches assigned by US Soccer if they meet the criteria. This process has undergone changes over recent years to simplify the process and

give more individuals a chance to move up in classification if they meet the requirements. In the US PRO assigns the matches for the top three divisions of men's professional soccer and the top division for women's professional soccer. An official does not have to be part of PRO or PRO2 to be assigned a match but must be a national referee with US Soccer to be a part of PRO or PRO2.

The process of training and developing referees is a time consuming and costly process, therefore the retainment of referees is extremely important. Understanding what causes officials to stop officiating early is a key to retention. Webb et al (2020) found that abuse was pushing officials away at the grassroots level. This is an issue that is occurring in several countries. The National Officiating survey 2017 found 87% of officials in the US suffered verbal abuse. Dawson et al (2021) found that with officials in France and Netherlands, the number of years officiating was correlated with the amount of verbal and physical abuse as well as the likelihood of quitting. They also found that the intention to quit was associated with suffering verbal abuse and intimidation. They concluded that the abuse being suffered by referees is leading to the decrease in number of active referees as noted by Giel & Breuer (2020). Slater (2019) and Webb (2021) explain that the loss of officials due to abuse and COVID-19 could have a negative impact on the quality of officials at the elite level over time. Giel & Breuer (2020) found that experiencing abuse during a match negatively predicted an official's intention to continue officiating. Along with this they found that referees who felt they were respected were more likely to continue refereeing. They also found the perception of support was an important predictor of intention to continue refereeing.

Choi & Chui (2017) found that South Korean officials perceived organizational support positively impacted on job satisfaction and career commitment which both negatively impacted turnover intention. Martínez et al (2021) found that aggression by players, coaches and fans increased the precursors to burnout of emotional exhaustion and cynicism in grassroots officials in Spain. Hong et al (2019) and Soriano et al (2019) found South Korean and Catalan officials, respectively, felt more intrinsically motivated when they felt they were supported by their association. Gorczynski & Webb (2021) outlined the pathway of research needed to adequately take care of the mental health of match officials. These studies all provide evidence for the importance of reducing referee abuse and providing organizational support. The benefit to all parties is to reduce the likelihood of an official suffering burnout,

leading to the decision to quit officiating and further contributing to the current rate of turnover and loss of match officials. It is important to evaluate the referee support structures currently in place to determine if they can adequately support match officials.

Webb et al (2021) examined the current referee development structure in the English Football Association. They interviewed 24 current referees or referees' "coaches". They found that while improvements had been made, there were still issues with recruitment of new officials. The interviewees recommended better utilization of social media to help with bringing attention to the opportunity presented by refereeing. They also recommended such recruitment focus not only on young children, but on adults and targeting students in college and university. Another recommendation is to give leeway to each county in how recruiting is done. It is possible that what works for one county may not work for another county due to the size and demographic differences between them.

Once an official is recruited, they enter the development pathway. Webb et al (2021) found that the feedback from the interviewees was mixed with regards to the effectiveness of the current pathway. They felt that the selection criteria for Centre of Referee Excellence (CORE) was murky. They stated that officials were unclear what it took to be selected into the CORE program. The interviewees also stated that if an official was not selected into the CORE program, then the opportunities for education were limited to their county and the quantity and quality of education varying greatly by county. Some officials reported satisfaction with the education provided by their county while others felt they did not receive enough educational opportunities. Additionally, they felt the need for consistent quality assessment to understand the areas that are needed for personal improvement.\

Along with lack of assessments they felt that the creation of a mentorship system would be helpful to developing referees. Currently the opportunities are limited due to limited funding. Along with mentorship, psychological and administrative support was felt to be lacking as officials progressed up the path outside of CORE. They felt they did not always know who to contact when certain issues arose, and they did not have a designated resource to turn to as they worked to cope with the more difficult matches. While the current development structure in England may have some areas for improvement, it also provides a good

foundation from which to build. The next step is understanding what a successful referee looks like so that the development structure can be formatted to find and develop those officials.

2.3.2 Characteristics of Successful Referees

Slack et al (2013) interviewed 15 English Premier League (EPL) referees in their effort to understand referee excellence. They performed semi-structured interviews designed to explore factors that EPL officials believed underpinned excellence in soccer officiating. They organized the 111 raw-data themes that emerged into 8 higher-order themes containing 24 lower-order themes. The higher-order themes were 1. Mental toughness attributes, 2. Support networks, 3. Effective game management qualities, 4. Multifaced pre-match preparation, 5. Performance-level enhancement, 6. Opportunities to thrive, 7. Personal characteristics, and 8. Superior physical components. This was the first investigation into what factors underpin successful referee performance. This work provides more information on a framework for guiding future research areas.

Pina et al (2019) attempted to parse out what aspects of a referee make them successful. They interviewed twenty-four individuals with expertise in soccer and refereeing. The individuals were asked a series of questions to try to elucidate what characteristics allow a top-level referee to excel. After analysing the data, they identified 3 major categories and 24 subcategories which were mentioned by more than 50% of the interviewees. The three major categories were: individual preparation, game preparation and game management. The sub-categories were sometimes related to multiples of the major categories. Individual preparation was related to individual characteristics such as physical fitness, hydration, individual learning style, commitment, and leadership. Game preparation was more related to knowledge of the players, crew communication (communication between the referee crew during the match), knowledge of team tactics and ability to stay cool under pressure. Game management was related to in-game aspects such as anticipation, game-reading, leadership, and experience. One significant piece of this work is that it displays the interconnectedness of different traits and how they can affect multiple aspects of referee excellence. This study is important as it sets the groundwork for designing ways to potentially differentiate officials of different skill levels. It creates a template for ways to potentially improve the training of

referees and it provides a guide for avenues of future research into referees and referee excellence.

Referees in Categories 1, 2, 3 of Portuguese soccer over the ten-year time span of 2009/10 season through to the 2018/19 season were examined by Mendes et al (2021). They found that length of time on the National Board, age, previous history of playing, and proximity to a dense population of other National Board officials were predictors of an official having a higher ranking. This can provide some information for recruitment, talent identification and talent development. One could potentially increase the likelihood of finding future successful officials by screening for someone who lives in an area with other National Board officials and played soccer for an extended period. Additionally, if an official lives in an area that is limited in the number of National Board officials but played for an extended period, the provision of increased opportunities for contact with National Board officials could offset the effect of the reduced density where they live. This also shows the importance of making sure officials get consistent match assignments to increase their level of experience on their development path.

The work by Pina et al (2019) relates to the work by Slack et al (2013). While they have different higher-order ways of categorizing the data there is overlap in the lower-order categories. Knowledge of team tactics, communication, physical fitness, game management, nutrition, experience, and teamwork were all mentioned as important parts of referee excellence by both studies. The higher-order theme of opportunities to thrive from Slack et al (2013) also relates to the work by Mendes et al (2021) showing that referees who were consistently working matches were more successful. Opportunities to thrive was broken down into personal domestic honours, international experience and extensive referee grounding. It builds on the idea that when referees are consistently challenged with matches and rewarded with higher calibre matches, they may grow to be more successful. The need for support structures to help referees grow and advance seen by Webb et al (2021) is supported by Slack et al (2013) and the higher-order theme of support networks and services being discussed by 86.6% of the officials interviewed. The higher-order theme of performance-level enhancement also coincides with desire for development pathways discussed by the officials in Webb et al (2021).
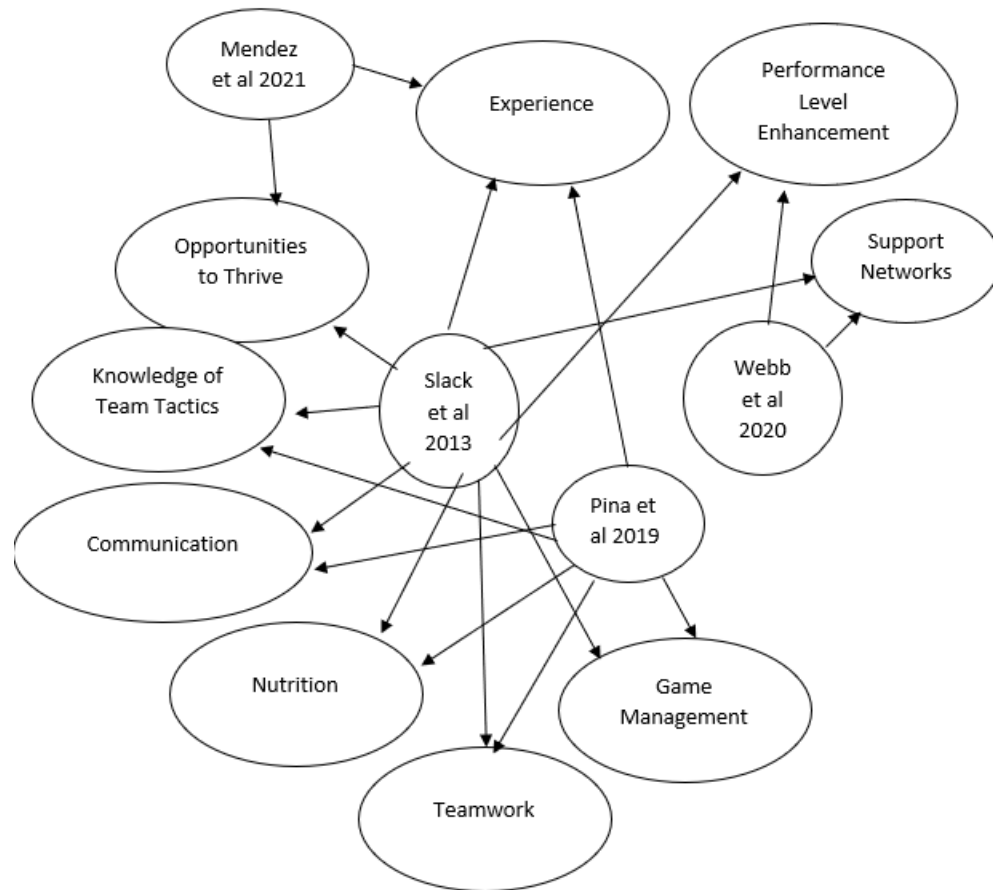
Figure 1 – Overlapping Characteristics of Successful Referees (Slack et al., 2013; Pina et al., 2019; Webb et al., 2020 and Mendez et al., 2021).

Loghmani et al (2021) interviewed twelve Iranian FIFA Referee and Assistant Referees to understand their life pathway in relation to their refereeing career. They divided the career into two stages, early (childhood, education periods, and development process) and late (elite pathways). In agreement with the work of Mendes et al (2021), they found that experience in the game of soccer before starting refereeing was connected with these officials who reached the elite level. Their finding of support from family and peers is coincides of the work by Webb et al (2021) and Slack et al (2013). One of their unique findings was the significant influence that each official's father played in the start of their careers. Another unique finding in the early period of the career is the importance of goal setting and individual practice to these officials successfully progressing up to the elite level. These can also be supported by the work of Slack et al (2013) with the lower-order theme of achievement striving (Highly motivated to be the best official, career focused).

They also found that enjoyment was important in the early period of the official's career, which as discussed above can be negatively impacted by several factors ultimately leading an official to stop officiating. Additionally, they discussed the need for an official to be willing to make sacrifices and miss family events and non-referee career opportunities early on in their career to gain the experience needed to advance in officiating. In the later period of their career, they found skill development, psychological state, minimizing the distraction by media, learning from mistakes, staying calm in matches, being supported by colleagues, minimizing bias during matches and a positive interaction with coaches and players as being important to success. These items are all in agreement with the previous work by Slack et al (2013) and Pina et al (2019).
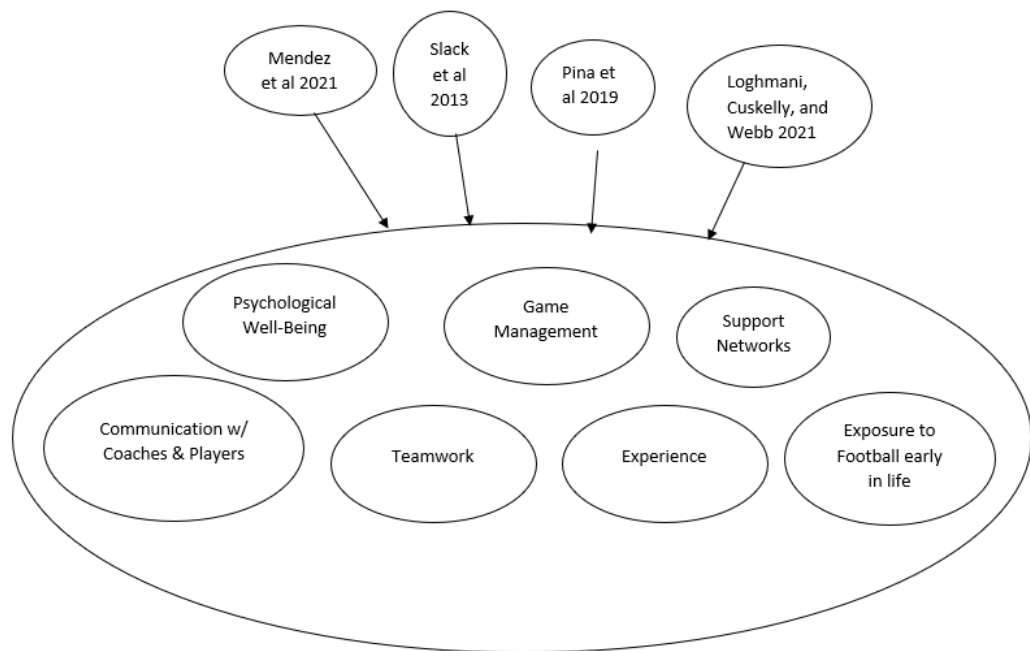
Figure 2 – Overlapping Characteristics of Successful Referees (Slack et al., 2013; Pina et al 2019; Loghmani et al., 2021 and Mendez et al., 2021).

2.2.3. Conclusion

The work of Slack et al. (2013), Pina et al. (2019)**,** Mendes et al. (2021) and Loghmani et al. (2021) each provide a different way of examining what defines a successful referee. The results of their work overlap in several areas which can been seen in Figure 2: experience, exposure to soccer from a young age, strong support network- both family and collegial, ability to manage the game, ability to interact with coaches and players, teamwork, and

psychological well-being. This work gives us an understanding of what makes a referee successful. The next steps are to take this work and build a way to evaluate officials to determine what areas they need to work on to become successful. There is yet to be anything done in this area in soccer referees. The aims of this project were 1. Conduct study 1 to determine what the definition of a successful referee performance is and what are the characteristics of a successful referee from a broad range of stakeholders in Major League Soccer (MLS). 2. Use the results of study 1 in collaboration with PRO2 coaches to develop a way to assess the referees: a) to differentiate between different levels of match official and b) to determine areas of strength and weakness for each match official. 3. Develop a battery of off-field methods to assess determinants of referee performance, based on items deemed important during the Delphi study which: a) have a strong correlation with the in-game assessment and b) differentiate between distinct levels of official.

Chapter 3

Determining the Definition and Components of a Successful Soccer Referee Performance

3.1 Introduction

A soccer referee is an athlete with a unique set of physical and mental demands during the performance of a match. The current duties of a referee according to the International Football Association Board (IFAB) are to "1. Enforce the laws of the game. 2. Control the match in cooperation with the other match officials. 3. Act as timekeeper, keep a record of the match and provide the appropriate authorities with a match report including information on disciplinary action and any other incidents that occurred before, during or after the match. 4. Supervise and/or indicate the restart of play." The financial impact of a referee properly performing these duties is increasing every year as the value of the teams and the prize money continues to climb. It is therefore important to understand what a successful referee performance is as well as which characteristics are possessed by a successful referee.

Research around referee success is relatively sparse with only four groups having examined it to date. Three groups performed interviews and one examined historical records of Portuguese National Referee Boards. Slack et al (2013) interviewed 15 English Premier League referees and organized the 111 raw-data themes that emerged into 8 higher-order themes containing 24 lower-order themes. The higher-order themes were: 1. Mental toughness attributes, 2. Support networks, 3. Effective game management qualities, 4. Multifaced pre-match preparation, 5. Performance-level enhancement, 6. Opportunities to thrive, 7. Personal characteristics, and 8. Superior physical components. This was the first investigation into what factors underpin a successful referee. Pina et al (2019) interviewed 24 individuals with experience in and around soccer. They developed 3 major categories and 24 subcategories. The 3 major categories were: Individual preparation, Game preparation, and Game management. The sub-categories were sometimes related to multiple of the major categories. Individual preparation was related to individual characteristics such as physical fitness, hydration, individual learning style, commitment, and leadership. Game preparation was more related to knowledge of the players, crew communication, knowledge of team tactics and ability to stay cool under pressure. Game management was related to in-game aspects such as anticipation, game-reading, leadership, and experience. One significant piece of this work is that it displayed the interconnectedness of different traits and how they can affect multiple aspects of referee excellence.

Loghmani et al (2021) interviewed 12 Iranian FIFA referees and assistant referees about their career development focusing on two time periods, the early development period and the elite pathway period. They found officials who were exposed to soccer prior to officiating and who had a strong support network were more successful later in their career. They found that goal setting and individual practice were important for an official to progress from the development period to the elite pathway. This was the first work looking at the entire career pathway of an official. Mendes et al (2021) found that the length of time on the Portuguese National Board, age, previous history of playing, and proximity to a dense population of other National Board officials were predictors of an official having a higher ranking. Overall, the work of each of the aforementioned groups provides a different way of examining what a successful referee is. The results of their studies overlap in several areas: Experience, exposure to football from a young age, strong support network both family and collegial, ability to manage the game, ability to interact with coaches and players, teamwork, and psychological well-being. These four works give us an understanding of what makes a referee successful for 4 different countries examining several different groups.

One area that is missing is the viewpoint on successful officiating from the league in which the officials work. There has been no consideration of the head coaches, general managers, or league officials' opinions regarding what attributes make a successful referee. There is also no clear definition of what it means for a referee to have a successful performance as determined by the stakeholders of refereeing. Most previous studies only asked referees for their definition of a successful soccer referee. Gathering the input of coaches, general managers, and league officials would further our understanding of what constitutes successful referee performance. Moreover, these few previous studies also only focused on a limited number of demands placed upon a referee, e.g., communication, interpersonal skills, etc., whereas the demands on a referee are multi-faceted across a range of areas, e.g., physical, technical, and psychological. Furthermore, these previous studies did not attempt to differentiate between skill levels of officials across these various demands. Such information is critical for development and promotion and would thus help referees and their coaches/managers to objectively identify the requirements for a referee to advance their career/level. Therefore, the aims of this study were to determine what the definition of successful referee

performance is and what are the characteristics of a successful referee from a broad range of stakeholders in MLS.

3.2 Methods

3.2.1 Participants

There were 38 respondents to part 1 and 52 respondents to part 2 of the survey. The officiating experience of respondents ranged from Youth, MLS Academy Exhibition games to FIFA World Cup games. There were 12 former or current FIFA Officials who responded. The coaching experience included National Team, MLS, USL, College, High School, and youth soccer. The playing experience included National Team Professional, College, Amateur and youth soccer. The breakdown of the respondents can be seen in Table 1.

Table 1 – Respondent's role within MLS and the number of respondents for each part.

Respondent	Role	38 Respondents in Part 1	52 Respondents in Part 2
MLS Coach	Head Coach in MLS	2	13
MLS General Manager	General Manager of an MLS team	6	12
MLS Officer	Works in the MLS office and has interactions around refereeing	1	2
PRO Assessor	Evaluates the performance of the referee during MLS matches	5	6
PRO Assistant Referee	Assistant Referees who work MLS matches	10	7
PRO Referee	Referees who work MLS Matches	4	3
PRO Staff	Works for PRO in Training or Coaching the Referees	8	5
PRO2 Referee	Primarily Referee 2nd Division Men's or 1st Division Women's matches, may be 4th official in MLS	2	4

3.2.2 Procedures

The Delphi study methodology was utilized to find a consensus on the definition of a successful referee performance and the items that make up a successful referee in MLS. The Delphi method is a way to solve a complex situation through the organization of individual opinions into a cohesive consensus (Linstone & Turoff, 1975). The typical process is to send a questionnaire or question to a select group of topic experts. The results are then used to formulate a second round of questions. This process is continued until the research question is answered. (Skulmoski et al., 2007) A two-part Delphi study was utilized to determine the

41

definition of a successful referee performance and the items which make up a successful referee.

After receiving local Ethical Approval (Reference# **19/SPS/053)** an email was sent out to the MLS Coaches and General Managers via an MLS Officer and to PRO Referees, Assistant Referees, Assessors, Staff and PRO 2 Referees by PRO Management. The email detailed the purpose of the project and provided a link to complete the questionnaire for part 1 online (Please see Appendix 1). The questionnaire remained open for Dec2019-Jan2020. All responses were anonymous with the goal of eliciting the most honest answers. The questionnaire contained two open-ended questions: Question 1. What is your definition of a successful referee performance? Question 2. What characteristics (e.g., mental toughness, composure, focus etc.) or abilities (e.g., law knowledge, positioning, sprinting speed etc.) make a referee successful?

Once Survey 1 was closed, the results were downloaded and analysed. The primary investigator grouped the responses by similarity. An item was created if there were three mentions of it in the results. The raw results were sent to two other investigators to group without having previously seen the primary investigators results. The other two investigators grouped responses in a similar manner to the primary investigator, confirming the formation of the items.

There were 7 items mentioned 3 or more times by respondents answering question 1. There were 26 items mentioned 3 or more times in response to Question 2. These items were then placed into Survey 2 (please see Appendix 2). The link to Survey 2 went out through the same channels as Survey 1. Participants were asked to rank the importance of the 7 items created by Question 1 and the 26 items created by Question 2 using a 7-point Likert Scale: 1 Not at all important, 2 Low Importance, 3 Slightly Important, 4 Neutral, 5 Moderately Important, 6 Very Important and 7 Extremely Important. Survey 2 stayed open for Feb20-Apr20.

3.2.3 Analyses

Once Survey 2 closed, the results were analysed and the 70% inclusion threshold was used for scores of 5,6, or 7 as these denoted positive importance value. Previous literature has used 70% consensus as the benchmark for inclusion. (Kleynen et al., 2014; van der Horst et al., 2017) All items above 70% agreement were included.

3.3 Results

The first goal was to establish a definition of what a successful referee performance is. The items in Table 2 were developed from the 89 total raw items mentioned by the respondents to Question 1 in Survey 1. Only items mentioned 3 or more times were included in Table 2 and in Survey 2. There were 15 items mentioned once which were discarded. All items included in Survey 2 achieved the 70% inclusion mark and were therefore included in Table 2 under Agreement %.

Table 2 – Reponses to Survey 1&2 Question 1 "What is your definition of a successful referee performance?"

A Referee Performance is Successful when	Mentions	Agreement %
No one talks about their impact on the game/they do not negatively impact the game	17	94.4
They make very few errors/all major decisions are correct	15	94.4
They control the game	10	77.3
They correctly enforce the laws of the game	9	100
They protect the health and safety of the players	7	98.1
They communicate with the players	3	96.2
Both teams feel like they received equal treatment	3	75.5

The second goal was to determine the characteristics of a successful referee. The items in Table 3 were developed from the 311 total raw items mentioned by the respondents to Question 2 in Survey 1. Only items mentioned 3 or more times were included in Table 3 and Survey 2. There were 4 items mentioned twice and 17 items mentioned once. All items included in Survey 2 achieved the 70% inclusion mark and were therefore included in Table 3.

Table 3 – Reponses to Survey 1&2 Question 2 "What characteristics (e.g., mental toughness, composure, focus etc.) or abilities (e.g., law knowledge, positioning, sprinting speed etc.) make a referee successful?"

Characteristics and Abilities of a Successful Referee	Mentions	Agreement %
Physical		
Physically Capable	29	98.1
Correct Positioning	20	98.1
Correct Movement	11	96.2
Proper Use of Body Language	6	98.1
Technical		
Knowledge of the Laws of the Game/ Ability to Apply the LOTG	28	100
Football Understanding	14	100
Strong Communicator	14	98.1
Ability to Manage Players/Coaches	9	96.2

Ability to Anticipate Play	7	96.2
Teamwork	7	94.3
Makes the Correct Decisions	4	98.1
Experience	3	77.4
Psychological		
Composed Under Pressure	26	100
Focus	20	100
Mentally Tough	20	98.1
Confident	8	98.1
Adaptable	4	96.2
Courageous	4	90.6
Ability to Lead	3	90.6
Strong Personality	3	80.8
Professional		
Approachable	12	96.2
Committed	9	94.3
Humble	8	86.8
Mentally Prepared	6	100
Trustworthy	6	94.3
Student Mentality	4	90.6

3.4 Discussion

The aims of this study were to obtain the opinions of key stakeholders within MLS surrounding the definition of a successful referee performance and the key characteristics and abilities of a successful referee. The main findings of the study were that seven items were identified in the definition of a successful referee performance which show that refereeing is a complex endeavour. In addition, 26 items were documented to comprise a successful referee, which reinforces the multi-disciplinary aspect of evaluating and training referees.

The definition of a successful referee as expressed by the participants here builds on the definition provided by IFAB (the body that determines the laws of the game of association football) by adding they communicate with the players and coaches, no one talks about their impact on the game/they don't negatively impact the game, they make very few errors/all major decisions are correct, they protect the health and safety of the players and both teams feel like they received equal treatment. This is the first work that attempted to define a successful referee performance.

The 26 items identified as what comprises a successful referee by the participants allows for a multidisciplinary examination of what it takes for a referee to be successful. This can allow

an up and coming official to breakdown their performances and understand the areas in which they need to improve upon in order to reach the next level. The four physical characteristics combine pure physical ability (Physical Capable) with correct utilization of their physical abilities (Correct Movement, Correct Position and Proper Use of Body Language). The eight technical characteristics balance the laws of the game with what is expected by football and what is needed to manage the players and coaches (Knowledge/Application of the Laws of the Game, Football Understanding, Player/Coach Management). The official works with their teammates to make sure the correct outcome is reached (Teamwork, Communication). These abilities are developed over many years as the official gains experience to help them read the game effectively and know where to be and what needs to be done (Ability to Anticipate Play, Experience, Makes Correct Decisions). The eight psychological characteristics are a combination of traits that help an official succeed during intense matches (Mentally Tough, Adaptable, Courageous, Composed Under Pressure, Focused, Confident) and keep control of the game (Strong Personality, Ability to Lead). The six professional characteristics are traits that lead to an official continually working to improve themselves before and after each match (Committed, Humble, Mentally Prepared, Student Mentality), and to be someone who others want to work with (Approachable, Trustworthy).

There are four other groups that have examined what makes a soccer referee successful. The work of Slack et al (2013) and Pina et al (2019) are similar in scope to this current project. These previous studies, as well as the current study, all asked the question what makes a referee successful and examined referees from the highest level of soccer in their respective country. There are commonalities in the findings of the three projects shown in Table 4. There were ten characteristics that were seen across all three studies: physically capable, correct positioning, experienced, mentally tough, mentally prepared, composed under pressure, football understanding, strong communicator, ability to manage players and coaches, and teamwork. This provides strong evidence that these common themes are important to a referee being successful. The present study differed from Slack et al (2013) and Pina et al (2019) in the use of anonymous survey as opposed to in person interview. . It is possible that different responses would have been found in the present study if in-person interviews were utilized, with more opportunity for discussion and clarification. However, the survey format allowed for a greater number of respondents as well as a more diverse set of respondents

who likely would not have had time for an in-person interview. In addition, potentially more honest answers were given due to the survey being anonymous.

One unique aspect of the current project was to ascertain the viewpoints of the coaches, general managers, and league officials from the MLS in which the referees work, in addition to referees, assistant referees, developmental referees, referee instruction staff and assessors. The goal in doing this was to attempt to provide a well-rounded perspective of successful officiating within the league, in order to most effectively screen and train referees for the league. Slack et al (2013) exclusively interviewed referees while Pina et al (2019) interviewed referees and assistant referees, referee coaches, referee committee members, practicing and academic sport psychologists, observers, journalists and physical and technical coaches. The inclusion of a broader set of respondents hopefully provides more support to the importance of these noted characteristics across a large number of stakeholders.

These three studies also differ in how they chose to organize their raw data into items, themes and categories. The present study combined the raw data into traits housed under four major categories focused specifically on the referee themselves. Slack et al (2013) chose to group raw data items into lower-order and higher-order themes. The 8 higher-order themes included referee specific themes, organizational and environmental themes and match focused themes. Pina et al (2019) organized their results into three dimensions with overlapping variables within the three dimensions. The dimensions, individual preparation, match preparation and game management were framed around the match as the focal point. Each method has its strengths and weaknesses. The present study isolates referee-specific variables in a way that easily allows for assessment and potential differentiation between skill levels. This removes aspects of interrelatedness as well as external factors that could affect referee success. Slack et al (2013) encompasses factors external to the referee that contribute to their ability to be successful while, also being more difficult to utilize the results to differentiate between levels of official. Pina et al (2019) displays the interrelatedness of certain variables between preparing for the match and managing the match, but also does not encompass organizational or environmental factors affecting referee success. It is also difficult to differentiate between officials based on this approach. The organization of the present study was structured to lead to the creation of an assessment tool. Despite the differences between the three studies on who they received responses from and how they organized the

responses, several common themes were evident across all 3 studies, which strengthens the importance of the areas in which they agree.

The group of Loghmani et al (2021), took a different approach to investigating successful referee performance by conducting interviews that examined the career pathways of successful referees from when they started refereeing up to their current level. While the approach was different there was still some overlap between what they found was important and the result of the present study. They found in the early period of a referee's career that goal-setting and individual practice were important to these officials successfully progressing up to elite level. This is related to the trait of a student mentality found in the present study. They also found that officials needed to be committed and make sacrifices early in their careers to develop into successful referees. When looking at the overlap between these early career traits found by Loghmani et al (2021) and the present study it is possible to see these traits being important to a successful referee's development process.

There was minimal overlap between the current work and the study of Mendes et al (2021). Their focus was narrower and more limited to descriptive data about referees as opposed to being able to ask deeper questions about them. Their results, obtained via examining historical records for Portuguese National Board officials over a ten-year period, provide an idea of how an official may develop and improve the traits that make them successful as opposed identifying the traits that make them successful. They found that proximity to other National Board officials, age, time on the National Referee board and playing history were positively related to an official's ranking. It is possible that by being in close contact officials can develop and improve the traits needed to be successful found in the present study. Time on the National Board and age allow for the official to develop the experience mentioned in the present study and develop traits such as football understanding and knowledge of the laws of the game. There may be minimal overlap in results between Mendes et al (2021). and the present study. However, they do complement each other with Mendes et al (2021), providing a way for the traits mentioned in the present study to be developed and improved.

Table 4 – Overlapping Traits of a Successful Referee from the results of the present and 2 previous studies

Trait	Study		
	Current Study	Slack et al (2013)	Pina et al (2019)
Physical Capable	x	X	X
Adaptable	x		X
Ability to Lead	x		X
Committed	x		X
Ability to Anticipate Play	x		X
Correct Positioning	x	X	X
Correct Movement	x		X
Student Mentality	x		X
Experience	x	X	X
Mentally Tough	x	X	X
Personality	x		X
Mentally Prepared	x	X	X
Composed under Pressure	x	X	X
Focus	x		X
Football Understanding	x	X	X
Strong Communicator	x	X	X
Ability to Manage Coaches and Players	x	X	X
Teamwork	x	X	X
Makes Correct Decisions	x	X	
Approachable	x	X	
Humble	x	X	
Honest	x	X	
Confident	x	X	
Nutrition		X	X

In the present study, the results to Question 2 coincide with the refereeing literature as well as when thinking about what a referee does when on the field. Most of these items are evident when watching a referee during a match. One example is the physical category. One can tell if they are physically capable of handling the demands of a particular match if their work rate decreases though out the duration of the match or if they are consistently left behind by the players even when they are able to anticipate play. The ability of an official to correctly move and to position themselves are two more items which can be evaluated easily during a match. In addition to the physical category, the technical category can be assessed during the match. An evaluator can assess the official's ability to make the correct decisions by watching the entirety of the match to make sure the decisions are correct within the context of the match and not just in isolation. When viewing the entire match, the assessor can evaluate the official's football understanding by observing how well the official's decisions fit the game.

Their use of personality, body language and ability to manage coaches and players can be identified during a match.

Some of the psychological categories can be evaluated during the match, for example, their composure under pressure during mass confrontations or when they need to make a big decision as well as their ability to focus for the entirety of the match and not miss small details. The mental toughness can be assessed by their ability to rebound after a mistake. It is important that these items be recognizable during the match so that they can be used to evaluate the performance of the referee to determine if they are ready for the next level or what areas they need to improve upon.

There are a few areas that are difficult to examine from an on-field performance, but which could be evaluated during a post-match debrief session. The evaluator could determine if they mentally prepared for the match and determine how well they take feedback i.e., if they have a student mentality to learn and grow or if they reject feedback. Their honesty could be assessed at this time as well, whether they are truthful when discussing what happened during the match. The postgame debrief can also be a time to assess if the official is confident yet humble, and if they are approachable. It is important that these items can be evaluated within the context of the match as ultimately that is where the referee performs.

The item of commitment can be evaluated by looking at the official's availability throughout the duration of the season. The experience of an official can be taken into consideration by looking at the number and quality of matches they have officiated throughout their career. It would be beneficial from a resource perspective if these items could be effectively evaluated off the field to screen potential candidates to work professional matches. The next steps of this project are two-fold. One step is to attempt to evaluate these items off the field to examine if a method can be created to differentiate between officials at different skill levels which would facilitate training/promotion without having to assess the official on the field. The other step is to develop a scoring system to assess the on-field performance of a referee using these characteristics and abilities as the basis.

3.5 Conclusion

This work was able to define what a successful referee performance means within MLS and what the characteristics of a successful MLS Referee are. There are seven items in the

definition of a successful referee performance showing that refereeing is a complex endeavour. These 26 characteristics and abilities indicate that evaluating and training referees is a multi-disciplinary endeavour. The commonalities with previous studies on successful referee give the essence of what aspects are important for soccer referees to be successful in different countries. It is important to attempt to find a way to assess these aspects when scouting future officials to determine if they have the characteristics necessary to be successful. More work needed to understand the effect of aspects that do not overlap and how they may influence a referee's success.

Chapter 4

Development of a Referee Game Assessment Program

4.1 Introduction

The value of professional soccer teams has been increasing year over year as the value of the players continues to increase. The top four leagues have seen an increase of 200% over a recent 13-year period (R Bloggers, 2020). The increasing value of the teams puts more value on each match. Soccer referees' decisions can have a major impact on the outcome on the match, influencing who wins and who loses. Soccer referees are unique athletes given the combination of physical and mental demands they are under during a match. Early research into refereeing began with trying to understand the physical demands (Weston, 2012; Weston et al., 2007; Weston et al., 2011; Castagna, 2004). It then progressed to try to understand the decision-making of referees by examining decision-making during the match and off-field using video-based methods (Helson and Butynek, 2004; Castillo et al., 2019; Hossner et al 2019; Mallo et al., 2012; Russel et al., 2020; Mascarenhas et al., 2009; Samuel et al., 2020; Ghasemi et al., 2011; Spitz et al., 2016; Morales et al., 2020). The focus has now moved to trying to understand what makes a referee successful as well as how to find and retain quality referees (Slack et al., 2013; Pina et al., 2019; Mendez et al 2021; Loghmani et al., 2021; Webb et al., 2020). Ultimately the goal of all this work is to improve soccer refereeing to keep up with the increasing demands of soccer. One attempt to improve refereeing is the establishment of companies with the specific goal of developing and managing soccer referees for professional matches.

The Professional Referee Organization (PRO) is a company in the United States which is contracted by Major League Soccer (MLS) to train and develop referees to officiate in that league. PRO2 is a branch of PRO that has two main objectives: 1. Train and develop referees who will be promoted to MLS. 2. Train and assign referees to officiate the National Women's Soccer League (NWSL) and two of the United Soccer League divisions, Championship (USL-C) and League 1 (USL-L1). There are three individuals who are responsible for coaching and evaluating the match officials for PRO2. They along with a video analyst and an administrator make up the PRO2 staff. The officials within PRO2 are divided into 4 primary tiers, A, B, C and D with a 5th tier being created for prospects. The potential readiness for MLS matches is A to P in descending order of readiness. Table 5 provides descriptions for each group of officials in this study. Understanding how to differentiate between different levels of match officials and where they need to improve is vitally important to the function of PRO.

Table 5 – Description and Roles for officials within PRO and PRO2

Tier	Description	Roles
MLS	Senior Match Officials	Officiate Matches for MLS
Tier A	Officials deemed most ready for trial in MLS	Officiate Matches for USL-C, NWSL, 4th official in MLS, Possibility of trial matches in MLS
Tier B	Officials deemed in training for a trial in MLS	Officiate Matches for USL-C, NWSL, 4th official in MLS
Tier C	Officials primarily focused on servicing USL and NWSL	Officiate Matches for USL-C, NWSL, USL-L1, 4th official in MLS
Tier D	Mix of Officials with potential for MLS and those that are primarily aimed at USL/NWSL	Officiate Matches for USL-L1, NWSL, USL-C
Tier P	New Prospects being evaluated	Officiate Matches for USL-L1, possible trial matches in USL-C, NWSL

Assessment of a referee's performance is crucial to examining the ability of elite referees as well as to facilitate the monitoring and progression of referees. There are currently very few, if any, methods to assess the multi-disciplinary demands of a referee during match play. The previous chapter divided the characteristics which make a referee successful into four major categories in order to facilitate the assessment of officials as to their readiness for MLS as determined by those with a stake in referee performance in MLS. These are Physical (physically capable, correct positioning, correct movement, strong body language) technical (knowledge of the laws of the game/ability to apply the laws of the game, football understanding, strong communicator, ability to manage players/coaches, ability to anticipate play, teamwork, makes the correct decisions, experience), psychological (composed under pressure, focus, mentally tough, confident, adaptable, courageous, ability to lead, strong personality) and professional (approachable, committed, humble, mentally prepared, trustworthy, student mentality).

Using these characteristics and categories, an in-game assessment metric can be developed which can differentiate between different levels of match officials, identify areas of strengths and weaknesses for each official and understand where best to apply the limited resources of time and money to the training of the officials. The purpose of this study was to therefore 1) to develop an in-game assessment instrument of referees based on the results obtained from the previous chapter, and in collaboration with members of PRO2, 2) examine the validity of the instrument by comparing between different levels of match officials.

4.2 Methodology

4.2.1 Design of an In-Game Assessment Instrument

The results of the Delphi study in Project 1 were given to the PRO2 staff. The PRO2 Staff consists of the Director of PRO2 Referees, two referee coaches who are responsible for coaching and evaluating the match officials, a video analyst, and an assignment coordinator/administrator. The Director and the two coaches are the individuals who watch and assess the matches. The PRO2 staff discussed the different items found in Project 1 and developed an assessment instrument. This was a collaboration of the scientific process and the expertise of these coaches. The result was named the Numerical Officiating Valuation Assessment (NOVA) system. The breakdown of the scoring process can be seen in Table 6 along with the Delphi item used to create each NOVA category and a brief definition of the category.

Table 6 – NOVA scoring system categories with definition, point value and Delphi item classification.

Major Category	NOVA Category	Delphi Item	Points	Definition
Technical	Misconduct Recognition	Makes correct decisions	10	Misconduct Recognition (MR) is the ability to successfully arbitrate situations that require misconduct designations.
	Foul Recognition		10	Foul Recognition (FR) is the consistent and credible recognition of foul play. Inconsistent foul selection can quickly derail a good performance.
	Match Control	Football Understanding	5	Match Control (MC) is the combination of foul/misconduct recognition/participant interaction applied into practice to keep a game in "cruise control." The smoother the game, the higher the match control score.
	Game Management		5	Game Management (GM) is the ability to foster an environment to promote the safety of players and participants and adequately maintaining time for stoppages and other in-game administrative duties
	Laws of the Game	Knowledge and Application of the Laws of the Game	5	Laws of the Game (LG) is the demonstrable understanding & consistent application of IFAB laws and procedures.
	Crew Teamwork	Teamwork, Strong Communicator	5	Crew Teamwork (TW) is the ability for the officiating crew to create cohesion and effectively work as a unit to provide the best service possible to the game of soccer
Physical	Positioning	Correct Positioning	5	Positioning (PS) is the ability to exhibit proper body orientation and positioning mechanics to maintain appropriate lines of sight while simultaneously demonstrating credible proximity to offences to "sell" decisions

	Stamina	Physically Capable	5	Stamina (ST) is the ability to exhibit appropriate levels of fitness throughout duration of game and ensure that quality of movement does not deteriorate before the final whistle.
	Explosiveness		5	Explosiveness (EX) is the ability to exhibit explosive movement, visible change of pace and acceleration, and ability to stay in-phase during attacks/counterattacks
	Active Zone	Correct Movement, Football Understanding	5	Active Zones (AZ) is the ability to combine proper understanding of football tactics with agility and lateral dexterity to stay out of active player zones.
Psychological	Anticipation	Anticipation, Football Understanding	5	Anticipation (AP) is the ability to exhibit desired "football instincts" that can combine movement and intuition to boost the quality of positioning
	Participant Interaction	Ability to Manage Players/Coaches, Strong Communicator, Strong Personality	10	Participant Interaction (PI) is integral to surviving in the professional game. Ability to use personality, communication, and mannerisms to proactively manage game temperature comprises ten total points.
	Composure	Composure, Body Language, Mentally Tough, Confident	5	Composure (CP) is the ability to "keep your cool" and maintain mental resiliency during key match situations to not allow adversity to negatively affect performance.
	Focus	Focused	5	Focus (FO) is the ability to maintain concentration throughout the duration of the game and not allow mental lapses to affect the outcome of the game
Professionalism	Match Difficulty		5	Match Difficulty (MD) is the level of difficulty of a match. A USL Championship Final on ESPN2 is not the same as a USL League One Intra-Squad Preseason Closed Door Scrimmage.
	Professionalism	Strong Leader, Approachable, Mentally Prepared	5	Professionalism (PF) is successfully performing duties and exhibiting leadership indicative of a professional referee.
	Coachability	Trustworthy, Student Mentality, Humble	5	Coachability (CO) is the ability to take criticism and apply instruction to foster an environment of self-improvement.

The item of physically capable was broken into 2 categories: stamina and explosiveness. This was done to attempt to elicit different strengths and/or weaknesses of the referees regarding the multi-faceted physical demands of refereeing a match, namely aerobic and anaerobic demands. Correct movement was focused on specifically in situations where they are close to play or potentially within passing lanes, these situations were deemed the "active zones." Proper use of body language was included in the category of composure as that is one of the

ways in which an official's composure can be displayed. The coaches felt it was important to differentiate between officials' ability to correctly recognize fouls and their ability to correctly recognize and administer misconduct. Experience was not something that the coaches felt would be an important piece of match evaluation, but it is considered during the end of the year evaluation process regarding promotion/relegation between tiers. Adaptable and courageous were not included as it was determined it would be difficult to evaluate during the match. Committed was not used in the game assessment but is used in the end of the year evaluation process regarding promotion/relegation between tiers. It is evaluated by availability and performance of off-field responsibilities. Three categories were given a value of 10 points because the coaches felt these were the most important categories and the others were given a value of 5 points so that the total would add up to 100 points.

4.2.2 Implementation of an In-Game Assessment Instrument

Figure 3 depicts the development of the NOVA instrument prior to its implementation for the duration of the 2021 USL-C season. All three coaches watched and assessed 4 USL-C matches by video from the 2020 season officiated by 4 different officials. They did this on two occasions. They met to discuss their results and gain consensus for each item of the NOVA instrument after watching all four matches the first time. They then re-watched and re-scored the same four matches and met and discussed their results again.
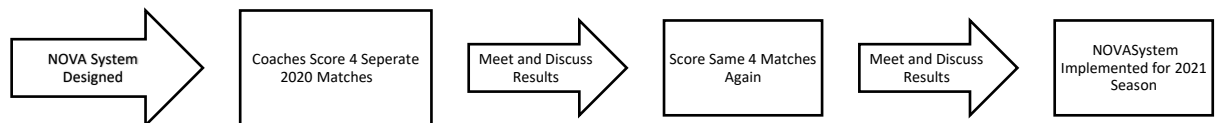
Figure 3 – Process for Design and Implementation of NOVA Scoring System

The intraclass correlation (ICC) was calculated via an ANOVA with two factor without replication was used for each variable and the total score of the NOVA and the following formula was used to calculate the intraclass correlation.

$$(MS_{Rows} - MS_{Error}) / (MS_{Rows} + df_{Columns} * MS_{Error} + (df_{Columns}+1) * (MS_{Columns} - MS_{Errors}) / (df_{Rows}+1))$$

MS – Mean Square
df – Degrees of Freedom

The NOVA system was then implemented for the 2021 USL-C season.

The USL-C is a second division league in the United States. The coaches evaluated the performance of 55 officials over 304 matches with the number of matches evaluated per official ranging from 1 to 12 matches. All matches were assessed via video with two different angles available (Broadcast and Wide View). The average and median number of matches evaluated per official was 5. The number of matches evaluated by each coach was 58, 116, 130, respectively. Coach 2 and Coach 3 are responsible for personally coaching 13 and 14 officials, respectively, within Tiers A, B, and C.

4.2.3 Analysis

A single factor analysis of variance (ANOVA) was run to determine between tier differences for each item of the NOVA system including Total Score. Significance was set to $p < 0.05$. If there was a significant difference between tiers a T-Test was run comparing each tier for a total of 10 comparisons. A Bonferroni correction was applied, with significance set at $p < 0.005$. A single factor ANOVA was run for between coach differences for each item in the NOVA system including total score. Significance was set to $p < 0.05$. If there was a significant difference between coaches a T-Test was run comparing each coach for a total of 3 comparisons. The Bonferroni correction was applied, and p was set to < 0.017. In order to examine if there was bias when a coach evaluated an official, they were coaching, an ANOVA was run for officials coached by Coach 2 comparing the 3 coaches scores. This was repeated for Coach 3 and the officials with No Coach. Significance was set to $p < 0.05$.

4.3 Results

The mean scores and the ICC values for each item can be seen below in Table 7. The ICC for Total Score for the first round was 0.11 and 0.98 for the second round. The intraclass correlation improved for almost every category from the first round to the second round as the coaches became more aligned with what they were looking for.

Table 7 – The means and intraclass correlation for each item and each coach for Round 1 and Round 2 of the design process.

	Round 1				Round 2			
	ICC	Coach 1	Coach 2	Coach 3	ICC	Coach 1	Coach 2	Coach 3
Misconduct Recognition	0.03	7.3±.6	8.3±0.3	9±2.0	1.00	7.3±4.3	7.3±4.3	7.3±4.3
Foul Recognition	-0.04	7.5±3.0	7.8±0.3	7±2.0	0.86	7.3±2.9	7.5±1.7	7.5±3.0

Participant Interaction	0.14	8.5±0.3	8±0.0	8.3±0.9	0.60	8.3±0.3	8.3±0.3	8.3±0.9
Match Control	-0.15	4.3±0.9	3.5±0.3	3.5±1.0	1.00	4.0±0.7	4.0±0.7	4.0±0.7
Game Management	-0.50	4.0±0.7	3.5±0.3	3.3±0.9	0.47	4.0±0.7	3.5±0.3	3.5±0.3
Laws of the Game	1.00	5.0±0.0	5.0±0.0	5.0±0.0	1.00	5.0±0.0	5.0±0.0	5.0±0.0
Teamwork	0.33	4.0±0.7	3.5±0.3	3.3±0.3	0.14	3.3±0.3	3.3±0.3	3.3±0.3
Anticipation	0.57	3.0±0.7	3.0±0.7	3.0±0.0	0.79	3.0±0.7	3.3±0.3	3.3±0.3
Positioning	0.08	3.3±0.3	2.8±0.9	3.0±0.0	0.41	3±0.0	2.8±0.3	3±0.0
Stamina	0.23	4.8±0.3	3.5±1.0	3.3±0.3	1.00	4.0±0.0	4.0±0.0	4.0±0.0
Explosiveness	0.11	4.3±0.3	3.3±0.9	3.0±0.0	0.73	3.5±0.3	3.5±0.3	3.3±0.3
Active Zone	-0.14	2.3±0.3	2.5±1.0	2.8±0.9	0.91	2.5±1.0	2.5±1.0	2.8±0.9
Composure	0.00	4.3±0.3	3.5±0.3	3.5±1.0	0.41	4.3±0.3	4.0±0.0	4.0±0.0
Focus	0.08	4.5±0.3	3.8±0.3	3.3±0.3	0.68	3.8±0.3	3.3±0.9	3.5±0.3
Match Difficulty	-0.14	4.0±0.7	2.8±0.3	3.5±0.3	1.00	4.5±0.3	4.5±0.3	4.5±0.3
Professionalism	1.00	5.0±0.0	5.0±0.0	5.0±0.0	1.00	5.0±0.0	5.0±0.0	5.0±0.0
Coachability	1.00	5.0±0.0	5.0±0.0	5.0±0.0	1.00	5.0±0.0	5.0±0.0	5.0±0.0
Total Score	0.11	81±44	75±39	75±34	0.98	78±47	77±44	77±41

The results for the between tier comparison can be seen in Table 8. There were eight items with significant differences between tiers. The key findings with regards to tiers were a significant difference between tiers for total score (0.001), specifically there was a difference between Tier A vs. Tier C (0.001), Tier D (0.001), Tier P (0.001) and between Tier B and Tier C (0.003), Tier D (0.001) and Tier P (0.001).

Table 8 – Means, Variance and Significance Values for the between tiers comparison

	Tier A	Tier B	Tier C	Tier D	Tier P	p-value
	43	92	56	98	15	
Misconduct Recognition	9.1 ± 2.0	9.3 ± 1.6	8.9 ± 3.3	8.7 ± 3.3	8.5 ± 2.0	0.078
Foul Recognition	7.5 ± 1.2 #	7.4 ± 1.0 #	7.4 ± 0.8 #	6.9 ± 1.2	7.1 ± 2.1	*0.001* *
Participant Interaction	8.1 ± 0.6 * # $	7.8 ± 0.6 #	7.6± 0.5	7.4 ± 0.4	7.3 ± 1.1	*0.001* *
Match Control	3.7 ± 0.3 # $	3.6 ± 0.3 # $	3.4 ± 0.4	3.3 ± 0.3	3 ± 0.1	*0.001* *
Game Management	3.1 ± 0.4	3.1 ± 0.3	3 ± 0.3	3.0 ± 0.21	3.0 ± 0.4	0.674
Laws of the Game	4.7 ± 0.5	4.8 ± 0.3	4.8 ± 0.3	4.7 ± 0.4	4.9 ± 0.3	0.734
Teamwork	3.1 ± 0.2	3.1 ± 0.2	3.0 ± 0.1	3.0 ± 0.34	2.9 ± 0.2	0.437
Anticipation	3.6 ± 0.3 * # $	3.3 ± 0.4 * #	3.0 ± 0.2	2.9 ± 0.2	3 ± 0.1	*0.001* *

					P-Values
Positioning	3.2 ± 0.4 **# $**	3.0 ± 0.3	3.0 ± 0.2	2.9 ± 0.4	2.7 ± 0.2	*0.002* *
Stamina	4.3 ± 0.4 **$**	4.4 ± 0.4 **$**	4.1 ± 0.6	4.3 ± 0.6 **$**	3.7 ± 0.7	*0.001* *
Explosiveness	3.7 ± 0.2 *** #**	3.7 ± 0.3 ***#**	3.2 ± 0.2	3.3 ± 0.3	3.5 ± 0.3	*0.001* *
Active Zone	3.1 ± 0.2	3.0 ± 0.2	2.9 ± 0.2	2.9 ± 0.2	2.9 ± 0.1	0.042 *
Composure	3.4 ± 0.3	3.3 ± 0.3	3.4 ± 0.3	3.2 ± 0.3	3.1 ± 0.2	0.039 *
Focus	3.2 ± 0.4	3.2 ± 0.4	3.1 ± 0.3	3.1 ± 0.4	3.1 ± 0.1	0.438
Match Difficulty	3.4 ± 0.3	3.3 ± 0.3	3.2 ± 0.2	3.2 ± 0.2	3.2 ± 0.2	0.104
Professionalism	4.9 ± 0.1	5.0 ± 0.01	5.0 ± 0	4.9± 0.2	5 ± 0	0.592
Coachability	5 ± 0	5.0 ± 0.1	5 ± 0	5 ± 0	5 ± 0	0.168
Total Score	77 ± 16 ***# $**	76 ± 15 ***# $**	74 ± 16	73 ± 19	72 ± 14	*0.001* *

Between Tier Post-hoc-BF 0.005 *** vs C, # vs D, $ vs P**

The results for the between coaches' comparisons can be seen in Table 9. There were items with significant differences between coaches. There were two key findings with regards to the between coaches' comparisons. The first finding was no significant difference between coaches for total score (P=0.65). The second finding is there was no significant difference in total score when the coach evaluates their own officials, Coach 2 (P=0.34) and Coach 3 (P=0.35) as seen in Table 10.

Table 9 – Means, Variance and Significance Values for the between coaches' comparison

	Coach 1 58	Coach 2 116	Coach 3 130	P-Values
Misconduct Recognition	8.4 ± 2.6 *	9.1 ± 2.3	9 ± 2.8	0.020*
Foul Recognition	7.6 ± 1.6 *	7.6 ± 0.6 ***	6.8 ± 1.2	0.001*
Participant Interaction	8.1 ± 0.4 * **	7.8 ± 0.4 ***	7.2 ± 0.6	0.001*
Match Control	3.2 ± 0.3 * **	3.5 ± 0.4	3.5 ± 0.3	0.004*
Game Management	2.8 ± 0.2 * **	3.0 ± 0.2	3.1 ± 0.3	0.003*
Laws of the Game	4.9 ± 0.1 **	4.8 ± 0.2 ***	4.6 ± 0.6	0.003*
Teamwork	3.1 ± 0.3	3.0 ± 0.2	2.3 ± 0.3	0.200
Anticipation	3.2 ± 0.4	3.1 ± 0.2	3.2 ± 0.4	0.508
Positioning	2.9 ± 0.2	2.9 ± 0.4	3.1 ± 0.3	0.069
Stamina	3.8 ± 0.3 **	3.8 ± 0.2 ***	4.9 ± 0.1	0.001*
Explosiveness	3.7 ± 0.4 * **	3.4 ± 0.2	3.4 ± 0.4	0.002*
Active Zone	3.0 ± 0.3	3.0 ± 0.1 ***	2.9 ± 0.2	0.034*
Composure	3.4 ± 0.3 *	3.1 ± 0.2 ***	3.4 ± 0.4	0.001*
Focus	3.4 ± 0.3 * **	3.0 ± 0.2 ***	3.2 ± 0.5	0.001*
Match Difficulty	3.5 ± 0.3 * **	3.2 ± 0.2	3.2 ± 0.3	0.001*
Professionalism	5 ± 0.1	4.9 ± 0.1	4.9 ± 0.2	0.982
Coachability	5 ± 0.1	5 ± 0	5 ± 0	0.336
Total Score	75 ± 23	74 ± 18	75 ± 20	0.654

Between Coach Post-hoc-BF 0.017. *Coach 1 vs Coach 2, **Coach 1 vs Coach 3, ***Coach 2 vs Coach 3

Table 10 - P-value means and variance between coaches for Total score for officials coached by Coach 2, Coach 3, and No Coach.

	Matches	Mean	SD
Coach 2		0.34	
Coach 1	19	75.26	33.32
Coach 2	35	75.71	14.62
Coach 3	33	74.06	23.25
Coach 3		0.35	
Coach 1	18	77.39	9.66
Coach 2	38	76.00	9.08
Coach 3	48	76.25	14.23
No Coach		0.35	
Coach 1	21	72.67	17.23
Coach 2	43	71.72	18.21
Coach 3	49	73.02	19.60

4.4 Discussion

The first aim of this study was to create an in-match assessment tool for soccer referees designed using the responses of individuals with a stake in refereeing in MLS for what they felt the characteristics of a successful referee were. The second aim of this study was to examine the validity of the instrument by comparing between different levels of match official. The two major findings from this project were strong inter-rater reliability for the total score and a significant difference between tiers with regards to the total score for the match. This is the first study that has been able to differentiate between skill levels of match officials based on an in-match assessment.

There are two characteristics evident in the literature which are able to differentiate between levels of referees: Decision making/foul recognition and anticipation. The differences between tiers for foul recognition in the present study are supported by previous work by Spitz et al (2016) who found that elite referees were more accurate compared to sub-elite referees during a video-based test using eye tracking software. This builds on the work from Gilis et al (2006) and MacMahon et al (2007) who found that referees were more accurate than non-referees (Players, Medical Staff and Coaches and Players) in recognizing fouls during video-based tests. It appears there is a specific skill set that is developed by referees for accurately recognizing foul situations. Catteeuw et al (2009) found that referees recognize fouls more accurately while assistant referees recognize offside more accurately.

It is possible that the skill set that allows referees to recognize fouls more accurately is an increased understanding of where to look and what to look for. Spitz et al (2016) found that elite referees spent less time fixating on irrelevant areas compared to sub-elite referees while also being more accurate during a video-based test using eye tracking software. Moreira et al (2020) found that national level referees had a shorter gaze fixation duration than state level referees during a video test. There were no differences in the number of gaze fixations or the accuracy of the decisions between levels. While gaze duration did not influence decision-making accuracy during the video test, it is possible that during a match the ability to pick up relevant cues in less time would lead to the more accurate foul recognition seen in the present study.

In addition to more accurate foul recognition a better ability to anticipate play was seen with higher tiers in the present study. This agrees with Spitz et al (2018) who found that elite referees performed better on domain-specific tasks of overall decision-making performance, anticipation, and recall capacity compared to sub-elite referees while they performed the same on domain generic tasks. Anticipation can play a role in not only allowing the official to be in the correct position to see what is occurring, but it is possible that anticipation has an influence on the combination of knowing where to look and more quickly being able to recognize the event occurring leading to more accurate decision-making on the field. This ability to anticipate play and recognize fouls correctly leads to a better performance and better total score for the match.

There has been very limited work designing and evaluating a performance assessment for sports officials. Jeffriess et al., (2020) examined National Rugby League referees to understand what factors affected performance evaluations. The model they used only explained 20% of the variance in score. According to their model penalty accuracy was the most valuable item for explaining the variance in scores. They found that referee rank did not affect performance score. Ekmekçi et al., (2020) attempted to improve upon the current basketball referee evaluation method in Turkey. They used a similar approach to the current study by asking the officials' opinion to help guide the weighting of the different variables. They felt this provided a more objective assessment of the officials. Neither of these studies attempted to differentiate between officials of different levels. Therefore, they were not able to examine what criteria differentiated one level of official from another.

The fact that Tier A and B officials scored significantly higher in total score than officials in Tiers C, D and P aligns with Tier A and B officials being the next in line for trial games in MLS. A trial game is when an official who is not a full-time MLS referee officiates an MLS match. The lack of a significant difference between Tier A and Tier B is not surprising as they have been selected with the idea that they may make it to the MLS within 2-4 years. In practice, the difference between Tier A and Tier B and the other tiers is the difference between a designation of a "Standard Performance" at 74 to a "Good Performance" at 76. It is a positive sign that Tier A and Tier B both average in the category of "Good Performance" with average total scores of 77 and 76, respectively.

The differences between the tiers can be instructive for ranking purposes as well as future education purposes. This can be illustrated well by examining the differences between Tier A and Tier D. Officials in Tier D scored significantly worse in anticipation, positioning, match control, participant interaction and foul recognition compared to Tier A. It is possible that these officials miss decisions due to poor anticipation of play and lack of explosiveness to get into the correct position to see the play. This, combined with a limited ability to interact with participants in a manner that lowers the temperature of the game, could lead to issues with match control. It is possible that a 3-pronged approach could help improve their game performance. The first prong is aimed at improving their ability to anticipate play. This could be targeted by discussions focusing on team tactics, it could also be targeted through more experience refereeing professional matches. The second prong could use physical training targeting their explosiveness abilities to attempt to allow them to make up for a reduced ability to anticipate play. The third prong would be developing more techniques to interact with players in a manner that would deescalate the temperature of the match and help the official to maintain control of the match.

The practical differences between the tiers can be helpful when discussing promotion/demotion of an official from one tier to another. It can be helpful to look at an official who outperforms the average of their tier and see where they stand in comparison to the other tiers. A season performance with an average of a tier higher could warrant a promotion to the higher tier. It also provides clear evidence when an official is not meeting the standard of their tier. If an official has an average performance placing them in the range of a lower tier it may merit a demotion. Ultimately it is up to those in charge to make the final

decision, but the clear and significant difference between tiers is helpful as a guide when making these decisions.

The strong inter-rater reliability is very important for the long-term implementation of this system. It allows for confidence from the officials that the score they receive from one coach is not different from the score they would receive from another coach. This is important as distrust in the coach's assessment can potentially hurt the learning process. The current inter-rater reliability also sets the standard for any additions to the coaching staff. There were significant differences in 12 of the 17 categories that contribute to the total score. It appears that Coach 3 potentially scores officials slightly lower on foul recognition and participant interaction and then buffers that by scoring stamina higher than the other coaches. The differences between coaches may reflect how each individual may focus on slightly different aspects while assessing a match, however these differences are not large enough to effect to total score.

In addition to inter-rater reliability, it is important that there is no expressed bias when a coach is evaluating an official with whom they work on an individual basis. There are three primary reasons that the coach would be evaluating their official: 1. It is important that they are able to watch them to provide quality advice on how to improve, 2. There are not enough PRO2 coaches to be able to evaluate all the matches without evaluating those they work with. 3. It is important that each coach evaluates each official as they each pick up on different aspects of the performance. The credibility of this assessment is strengthened by there being no significant difference between the scores for the officials when their personal coach is evaluating them and when the other coaches are evaluating them.

One difficulty in the implementation of this assessment process is it is very time intensive. In addition to watching the match and filling out the assessment the coach needs to have a debrief session with the crew. There were 514 matches during the 2021 USL-Championship season. That is an average of 18 matches per week over the 28-week season. The three coaches evaluated 304 matches during the season.

The next steps are to continue to utilize the assessment for the 2022 season and review for changes and to repeat the process to develop a way to assess Assistant Referees in a similar manner.

4.5 Conclusion

This study showed it was possible to differentiate between different levels of soccer referee using an in-game assessment designed with input from those involved in soccer at the highest level. This process could be implemented in any organization in the world following the same steps. This will allow a group to properly rank their officials for promotion/relegation using the total score as well as to understand how to help the officials improve by identifying where they are weak compared to their fellow referees. Ultimately, this project was successful in its aim to collaborate with the referee coaches of PRO2 to develop a way to assess the referees based on the results of study 1.

Chapter 5

Off-Field Assessment of Referee Characteristics as developed from a Delphi Study

5.1 Introduction

The Professional Referee Organization (PRO) is a company in the United States which is contracted by Major League Soccer (MLS) to train and develop referees to officiate in that league. PRO2 is a branch of PRO which has two main objectives: 1. Train and develop referees who will be promoted to MLS. 2. Train and assign referees to officiate the National Women's Soccer League (NWSL) and two of the United Soccer League divisions, Championship (USL-C) and League 1 (USL-L1). The officials within PRO2 are divided into 4 primary tiers, A, B, C and D with a 5th tier (P) being created for prospects. The potential readiness for MLS matches is A to P in descending order of readiness.

Understanding how to differentiate between different levels of match officials and where they need to improve is vitally important to the function of PRO. This led to the development of the Numerical Officiating Valuation Assessment (NOVA) instrument as described in the previous chapter that facilitates the assessment of the multi-faceted performance of a referee during a game. A Delphi study was used to ask those with a stake in refereeing in MLS what makes a referee successful. Multiple viewpoints were obtained to make sure all potential factors were accounted for as team coaches/general managers see different characteristics than current referees or referee assessors. The NOVA instrument was then used for the duration of the 2021 USL-C season and was able to detect differences between tiers for foul recognition, participant interaction, match control, anticipation, positioning, stamina, explosiveness and total score.

Although the NOVA instrument can provide valid and sensitive data on a referee's performance during a game, it is a time consuming and costly process requiring at least 1 PRO2 coach to observe and score the referee during the   match. Instruments that also assess key determinants of a referee's performance outside of a game setting are important to provide regular and objective monitoring of a referee's development for training in preparation for games and/or for promotion to higher tiers.

There have been a few studies that have attempted to differentiate between skills levels of referees off the field. The majority of the work has used video clip-based tests to differentiate between non-referees (players, coaches, medical staff) and referees (Gillis et al., 2006; MacMahon et al., 2007), elite vs. sub elite referees (Spitz et al., 2016) and referees

vs. assistant referees (Catteeuw et al., 2009). Referees were better able to recognize fouls compared to non-referees and assistant referees while elite referees were more accurate than sub-elite referees. Additionally, Spitz et al. (2018) found that elite referees were better able to anticipate the correct outcome of a potential foul situation when the video was shut off prior to the incident occurring compared to sub-elite referees. Such methods of assessment of referees are useful in that they directly assess foul recognition/laws of the game. However, as has been outlined in Chapter 3, a referee's performance involves a variety of inter-disciplinary demands such as physical, psychological, technical and interpersonal skills. Any off-field or remote assessments of referees should therefore encapsulate examinations of the range of demands placed on a referee. The goal of this study was to 1) develop a battery of off-field methods to assess determinants of referee performance based on items from the Delphi study and compare it to the NOVA, and 2) differentiate between different levels of official using the battery.

5.2 Methods

5.2.1 Participants

The participants were from three groups: PRO Officials, PRO2 Officials or Non-PRO Officials. The descriptions of each group can be seen in Table 11. The criteria for participation were being assigned a minimum of one match by PRO or PRO2.

Table 11 – Description and Roles for officials within PRO and PRO2

Group	Tier	Description	Roles
PRO	FIFA	Senior Match Officials who also is on FIFA Panel	Officiate International Matches assigned by CONCACAF and FIFA and MLS matches
	MLS	Senior Match Officials	Officiate Matches for MLS
PRO2	Tier A	Officials deemed most ready for trial in MLS	Officiate Matches for USL-C, NWSL, 4th official in MLS, Possibility of trial matches in MLS
	Tier B	Officials deemed in training for a trial in MLS	Officiate Matches for USL-C, NWSL, 4th official in MLS

	Tier C	Officials primarily focused on servicing USL and NWSL	Officiate Matches for USL-C, NWSL, USL-L1, 4th official in MLS
	Tier D	Mix of Officials with potential for MLS and those that are primarily aimed at USL/NWSL	Officiate Matches for USL-L1, NWSL, USL-C
Non-PRO Official		National or Regional Officials not part of the PRO2 program	Officiate Matches for USL-L1, possible trial matches in USL-C, NWSL

5.2.2 Design

In an earlier chapter the characteristics which make a referee successful were determined by those with a stake in referee performance in MLS and subsequently divided into four major categories based on the multi-faceted nature of soccer refereeing. Physical (physically capable, correct positioning, correct movement, strong body language) technical (knowledge of the laws of the game/ability to apply the laws of the game, football understanding, strong communicator, ability to manage players/coaches, ability to anticipate play, teamwork, makes the correct decisions, experience), psychological (composed under pressure, focus, mentally tough, confident, adaptable, courageous, ability to lead, strong personality) and professional (approachable, committed, humble, mentally prepared, trustworthy, student mentality). This study will examine each item in the psychological and professional categories as well as stamina, knowledge of laws of the game, communication and teamwork.

Psychological and Professional Components

A test battery was designed to examine the psychological and professional components of a successful referee. An attempt was made to find a survey that targeted each topic category. A search of the literature pertaining to each topic was performed and colleagues with experience in the field of sport psychology were consulted. There were six surveys found that appeared to be relevant to the items in the psychological category: TOPS 2-S (composure and focus), Vealey's Trait Confidence Questionnaire (confidence), MTQ-18 (resilience), BIA/BIS (courage), Measure of Transformational Leadership (leadership), and the Big 5 Personality Questionnaire (strong personality). The research team decided to design a

single question for each item in the professional category (approachable, committed, humble, mentally prepared, trustworthy, student mentality).

Once the entire battery was compiled, the research team decided that modifications were needed to reduce the length of the survey and improve the likelihood officials would be willing to complete it. The different tools were modified to remove repeat or irrelevant questions. The primary researcher selected the questions for removal and then the team reviewed them to confirm changes. The Likert scales for the TOPS 2-S, Vealey's Trait Confidence Questionnaire, BIA/BIS and the Measure of Transformational Leadership were also all adjusted to the same 5-point scale: 1. Strongly Disagree, 2. Disagree, 3. Neither Disagree nor Agree, 4. Agree, 5. Strongly Agree. This was done to minimize confusion when participants were completing the survey.

The survey was sent out anonymously via a member of PRO Staff to all officials who were assigned a match by PRO at MLS, NWSL, USL-C or USL-League 1.

Laws of the Game

A law of the game test was designed with the help of the PRO Head of Coaching, Education and Evaluation who provided resources that they had previously utilized to test officials in different contexts on the laws of the game. The primary researcher used these resources to develop a law of the game test. The survey was sent out to all PRO2 officials by a member of the PRO2 staff.

Communication and Teamwork

A survey was sent out to PRO2 officials with a series of questions investigating on-field communication. In the final part of the survey, they were asked to rate all officials in tiers A-D based on their ability to utilize effective communication and teamwork during a match. A 5-point Likert Scale was used 1 – Very Poor, 2 – Poor, 3 – Fair, 4 - Good, 5 - Excellent. A sixth option to select "have not worked with" was included.

Endurance Ability

2000m Time Trial

Approximately every three months officials in PRO and PRO2 are instructed to run a 2000m time trial. The Pro Director of Sport Science at the time decided that 2000m was the distance to be used. Given that the officials are distributed throughout the country these time trials are run alone and unsupervised. Officials are instructed to complete 2000m as quickly as possible and record their time using a watch.

The fastest time trial in the past 12 months for each individual was selected and used for analysis.

4-minute Submaximal Test

Officials were assigned a 4-minute submaximal treadmill assessment at various points throughout 2019 – 2021 as part of an ongoing monitoring program started by the Pro Director of Sport Science at the time. They were instructed to follow the procedure outlined in Figure 4. The heart rate was recorded using a Polar Heart Rate monitor distributed to them over the 2019-2021 seasons.
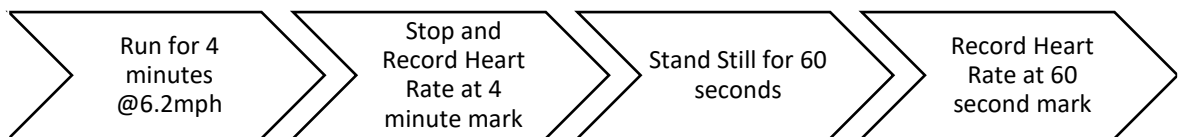
| Run for 4 minutes @6.2mph | Stop and Record Heart Rate at 4 minute mark | Stand Still for 60 seconds | Record Heart Rate at 60 second mark |

Figure 4 – Procedure for 4-minute Treadmill Submaximal Assessment.

5.2.3 Analysis

Psychological and Professional Components

A single factor ANOVA was performed between tiers for the TOPS 2-S (composure and focus), Vealey's Trait Confidence Questionnaire (confidence), MTQ-18(resilience), BIA/BIS (courage), Measure of Transformational Leadership (leadership), four of the components of the Big 5 Personality Questionnaire and each of the 7 questions in the professional category. Significance was set to $P<0.05$.

Laws of the Game

A single factor ANOVA was performed between tiers.

A correlation was run between the laws of the game test score and the laws of the game score for the 12 officials who had their matches assessed using the NOVA system.

Communication and Teamwork

A single factor ANOVA was performed between tiers. A series of T-tests were performed to examine the differences between tiers. A Bonferroni post-hoc test was used, and significance was set to $p \leq 0.008$.

A correlation was run between the peer rating of communication and the NOVA score for Communication.

Endurance Ability

A single factor ANOVA was performed between tiers for both the 2000m time trial and the 4-minute treadmill assessment. Significance was set at $p \leq 0.05$.

A t-test was performed to examine the differences between the referees and assistant referees for both the 2000m time trial and the 4-minute assessment. Significance was set at $p \leq 0.05$.

There were 30 referees who completed the 2000m time trial and the NOVA score for Stamina.

5.3 Results

Psychological and Professional Components

There were 108 respondents. After 11 were removed due to missing answers, there were 97 fully completed surveys over a 30-day period between January and February 2020. The breakdown by position was even with 53 referees and 54 assistant referees. The breakdown by level was 12 FIFA officials, 18 MLS officials, 57 PRO2 officials and 10 non-PRO2 officials

No difference was found between tiers for any of the questionnaires or questions. A non-statistically significant difference was found for two categories, composure and focus, and adaptable. The FIFA group scored approximately 6 points higher than the non-PRO2 referee group (Table 12).

Table 12 – P-Value, Mean and Variance for each category targeted by a questionnaire or question.

Category	p-value	FIFA AU 12	MLS AU 18	PRO2 AU 57	Non-PRO2 AU 10
Composure and Focus	0.19	82 ± 38	82 ± 23	78 ± 85	77 ± 51
Confidence	0.79	37 ± 18	36 ± 15	35 ± 26	35 ± 31
Resilience	0.70	29 ± 13	30 ± 14	30 ± 10	30 ± 13
Courage	0.37	32 ± 17	34 ± 25	34 ± 22	33 ± 40
Leadership	0.63	29 ± 2	29 ± 2	28 ± 5	28 ± 7
Total Extroversion	0.82	16 ± 41	17 ± 23	18 ± 23	17 ± 15
Total Agreeableness	0.49	30 ± 13	28 ± 22	29 ± 15	31 ± 19
Total Conscientiousness	0.30	13 ± 4	12 ± 5	12 ± 4	13 ± 6
Total Neuroticism	0.29	21 ± 19	18 ± 18	19 ± 20	21 ± 30
Reliable	0.86	4.9 ± 0.1	4.8 ± 0.2	4.9 ± 0.1	4.8 ± 0.2
Committed	0.54	4.7 ± 0.2	4.4 ± 0.7	4.5 ± 0.5	4.7 ± 0.5
Student Mentality	0.40	4.8 ± 0.2	4.7 ± 0.2	4.8 ± 0.1	4.8 ± 0.2
Approachable	0.65	4.3 ± 1.2	4.6 ± 0.5	4.5 ± 0.5	4.7 ± 0.5
Humble	0.81	4.8 ± 0.2	4.7 ± 0.2	4.7 ± 04	4.8 ± 0.2
Mentally Prepared	0.22	4.7 ± 0.2	4.8 ± 0.2	4.3 ± 0.9	4.4 ± 0.8
Adaptable	0.10	3.7 ± 2.0	4.4 ± 0.6	4.1 ± 0.8	4.7 ± 0.3

Laws of the Game

There were 33 responses: Tier A-2, Tier B-4, Tier C-8, Tier D-19. There was no difference between tiers. The mean score and variance for each tier can be seen below in Table 13.

Table 13 – P-value for ANOVA between tiers and mean score and variance for each for laws of the game test.

Laws of the Game Test			
Tier	Count	Mean Test Score	p-value
Tier A	2	28.5 ± 4.5	
Tier B	4	31.5 ± 16.3	0.074
Tier C	8	28.5 ± 20	
Tier D	19	25.2±24.4	

The correlation coefficient between the laws of the game test and the NOVA score for laws of the game was 0.47 meaning a moderate positive correlation between scores.

Communication and Teamwork

There were 16 respondents, 8 referees and 8 assistant referees. The breakdown between tiers was 1 tier A, 3 tier B, 2 tier C and 10 tier D.

There was a difference between tiers ($p = 0.01$). The difference between tier B and tier D was borderline significant (4.3 ± 0.2 vs 3.7 ± 0.5 $p = 0.008$). The correlation between the Peer Rating of Communication and NOVA score for communication and teamwork was low at 0.32.

Endurance Ability

There were 126 individuals who completed at least one 2000m time trial between 2020 and 2021. The time trial was completed by 47 MLS officials, 13 PRO2 Tier A, 16 PRO2 Tier B, 21 PRO2 Tier C and 29 PRO2 Tier D officials. There was no difference between the tiers or the positions for 2000m time $p = 0.051$ and $p = 0.65$, respectively. The mean times can be seen in Tables 14 and 15.

Table 14 – Mean 2000m time for each tier.

Tier	Count	Mean 2k Time (s)
MLS	47	481 ± 38
A	13	466 ± 53
B	16	476 ± 46
C	21	515 ± 88
D	29	516 ± 102

Table 15 – Mean 2000m time for position.

Position	Count	Mean 2k Time (s)
Referee	57	489 ± 75
AR	69	495 ± 68

There were 677 and 671 4-minute submaximal treadmill tests performed by referees and assistant referees, respectively. There were 709 tests performed by MLS officials, 148 by tier A, 135 by tier B, 141 by tier C and 215 by tier D.

There was a difference between tiers for heart rate after 4min (p = 0.001), heart rate after 1 minute standing recovery (p = 0.001) and recovery heart rate (p = 0.001). The heart rate at the end of the 4-minute run was lower for MLS officials, 135±16, compared to tier A, 143±18, (0.001), tier B, 146±16, (0.001), tier C, 147±13, (0.001) and tier D, 146±18, (0.001). The heart rate at the end of the one-minute standing recovery was lower for MLS officials, 97±18, compared to tier A, 106±16, (0.001), tier B, 108±17, (0.001), tier C, 111±21, (0.001) and tier D, 113±23, (0.001). The heart rate at the end of the one-minute standing recovery was lower for tier A compared to tier D 106±16 vs 113±23 (0.001). Recovery Heart rate was greater for MLS officials than tier D officials 38±12 vs 33±17 (0.001). The mean heart rates for each tier at each time point can be seen in Table 16.

Table 16 – Mean and standard deviation heart rate in beats per minute (bpm) for each tier at post 4-minute run, post 1 minute recovery and the recovery heart rate.

Group	Count	HR_Post 4min (bpm)	HR_Post 1 min(bpm)	Recovery (bpm)
MLS	709	135±16	97±18	38±12
A	148	143±18*	106±16*	37±13
B	135	146±16*	108±17*	38±12
C	141	147±13*	111±21*	36±14
D	215	146±18*	113±23*$	33±17*
	p-value	0.001	0.001	0.001

*Significantly different from MLS

$Significantly different from Tier A

The heart rate after 1 minute recovery was higher for referees compared to assistant referees 104±20 vs 102±20 (0.03) and the recovery heart rate was larger for assistant referees compared to referees 35±13 vs 39±14 (0.01). The heart rates for each position can be seen in Table 17.

Table 17 – Mean and standard deviation for heart rate in beats per minute (bpm) for referees and assistant referees at post 4-minute run, post 1-minute recovery and the recovery heart rate.

Group	Count	HR_Post 4min (bpm)	HR_Post 1min (bpm)	Recovery (bpm)
Referee	677	140±19	104±20	35±13
AR	671	140±16	102±20*	39±14*
	p-value	0.32	0.03	0.01

Significantly different from Referees

The correlation between the NOVA and the 2km time trial was very weak -0.1.

5.4 Discussion

The goal of this study was to develop a battery of off-field methods to assess determinants of referee performance based on the items deemed important during the Delphi which: a) have a strong correlation with the in-game assessment and b) differentiate between distinct levels of official. Regarding goal b, this study was able to differentiate between level of official based on a 4-minute submaximal treadmill test. It was unable to differentiate between levels of official for psychological and professional components, knowledge of laws of the game, and communication and teamwork. Regarding goal a, this study found a moderate correlation between the laws of the game test score and the laws of the game score for the NOVA project. There was a weak correlation between peer rating of communication and teamwork and the NOVA Score for communication and teamwork.

Unlike this study, Gillis et al. (2006), MacMahon et al. (2007), Catteeuw et al. (2009) Spitz et al. (2016), Spitz et al. (2018) were each able to differentiate between levels of official, Assistant Referee vs. Referee, or officials vs. non officials for decision making/foul recognition. They have shown that it is possible to differentiate between levels of official off the field using specific video assessments of decision making and anticipation. Such aspects were not assessed in the current study's battery. It is important to try to replicate this differentiation in other important aspects of officiating. The ability to differentiate between level of official off the field has the potential to improve resource allocation and improve monitoring of officials to determine when they are ready to step onto the field for a particular level of match.

Psychological and Professional Components

The purpose of this study was to attempt to differentiate between different levels of soccer referee's psychological and professional qualities using a series of questionnaires. This was not accomplished using this set of questionnaires. There is some agreement with the NOVA on-field assessment as there was no difference between tiers for the categories composed of psychological and professional components except for participant interaction (strong personality) and composure (composure, mentally tough, and confident). It may be of value to design more targeted assessments in order to differentiate between tiers for these specific components which were different between tiers for an on-field assessment. It is also possible that different results would have been seen had each questionnaire been administered in full on separate occasions as opposed to modified and in one large battery. One other possibility is that officials who make it to the professional level possess similar traits and the differences may be found between referees and non-referees.

Laws of the Game

The goal of this study was to attempt to differentiate between tiers of PRO2 officials based on a test of their knowledge of laws of the game. There were no significant differences between tiers. This agrees with the Laws of the Game score for the NOVA game assessment. There was no difference between tiers. It is possible that knowledge of the laws is equivalent between officials once they reach the professional level. There were some practical differences with Tier B scoring 6 points higher than Tier D on the test. The small number of officials participating in the higher tiers does leave room to question if there would be different results seen with more participation from officials at a higher tier.

The moderate positive correlation between the laws of the game score and the NOVA score for laws of the game is promising. There were only 12 officials who completed the laws of the game test and were evaluated via the NOVA, making it difficult to trust the strength of this correlation. A strong correlation between the two scores would lend more credibility to being able to use a law of the game score as a screening method. It may not be a way to differentiate between tiers but if is strongly correlated with on field performance then it could be used to screen candidates for entry into PRO2.

Communication and Teamwork

The aim of this study was to compare an off-field method for assessing communication and teamwork with the on-field method. There was a weak positive correlation between the on-field and off-field methods. It is unlikely that the off-field method would be a good proxy for the on-field method. Interestingly there was a difference between tier B and tier D with the off-field method while there was no difference between tiers for the on-field method. It is possible that the on-field method does not accurately pick up differences in communication and teamwork as the person evaluating the match cannot listen in to the communication, so they are on evaluating based on what they see and then what is discussed in the debrief.

One limitation to the teamwork element of this study is that 80% of the officials being rated received at least 8 responses of "Have not worked with". This means that at least half of the 16 respondents did not work with 80% of the listed officials. It is possible that the results of the survey would have been different if more officials had worked with the listed officials.

Endurance Ability

The purpose of this study was to attempt to utilize assessments which were already in place to differentiate between level of match officials with regards to the endurance ability. The 2000m time trial was not able to differentiate between the tiers. This is likely due to the variation of effort put in by officials, not all officials put in the best effort. The 4-minute was able to differentiate between the MLS level and the lower levels for heart rate post 4-minute run and post 1-minute recovery. Unfortunately, the MLS officials are not scored on the NOVA assessment so it cannot be compared and the only difference between tiers for stamina was between Tiers A, B and D and Tier P who did not participate in this study.

While not significantly different, there was a subtle difference between the MLS, tiers A and B from tiers C and D. The MLS officials completed the 2000m trial ~34 seconds faster, tier A ~50 seconds faster and tier B ~40 seconds faster than the officials in tiers C and D. The tier A and B officials having slightly faster times than the MLS officials makes sense as these are the up-and-coming officials who are trying to break through to the MLS level and therefore, pushing harder. The difference between tiers A and B and tiers C and D would be expected as the former are the officials who are expected to make the step up to the MLS level in the near future, while the latter are either prospects needing more time to develop or officials with a ceiling in the lower divisions.

The difference between the heart rate after the 4-minute run between the MLS and the lower tiers is possibly due to increased aerobic fitness level, or it is also possible that the MLS officials are more experienced with performing this test. The difference in recovery heart rate being ~ 0 beats for everyone except for tier D lends credence to the MLS and higher tiers having a stronger fitness level, but the non-MLS officials not being as practiced at the test, so their 4-minute heart rate is higher.

It appears that the 4-minute treadmill test used to assess match officials can differentiate between levels. It did differentiate between the MLS officials and all lower tiers for both the post 4-minute and post recovery heart rate and MLS and tier D for the recovery heart rate. However, it is potentially due to better familiarity with the test. While the 2000m time trial was not significantly different, it did appear to have clear practical differences in completion time between MLS, tiers A and B and tiers C and D. These tests have the potential to work as screens for endurance ability.

5.5 Conclusion

The goal of this study was to differentiate between levels of official based on the results of the Delphi project using off-field methods. Overall, this was unsuccessful. This study found no differences between the tiers for any of the psychological or professional qualities. Currently based on the results of this study the utilization of a laws of the game score does not work to differentiate between tiers of official nor does it work as a proxy for a match assessment score for knowledge of the laws of the game. Differentiating between level of official based on the feedback from other officials did not find agreement between the on-field and off-field measure of communication and teamwork. There was a difference between tier B and tier D, however, it is unclear if that difference would remain with participation from more officials. It appears that the 4-minute treadmill test used to assess match officials can differentiate between levels, but there are questions of familiarity being the cause for the difference. Ultimately, more work needs to be done to determine if differences in skill level for these criteria can only be found on the field.

Chapter 6

Synthesis and Reflection

6.1 Synthesis

Soccer referees' decisions can have a major impact on the outcome on the match, influencing who wins and who loses. Soccer referees are unique athletes given the combination of physical and mental demands they are under during a match. The majority of the referee literature is focused on physical aspects of officiating and the number of decisions made during a match. The research attempting to understand what makes a referee successful is relatively sparse with only four groups attempting to answer those questions. It is a vital question to those who work to develop and promote referees to officiate higher levels of competition.

6.2 The Problem

The mission of PRO is to train and prepare officials for MLS matches. The mission of PRO2 is to find and develop officials for promotion into PRO while also supplying officials for USL and NWSL matches. According to the US Soccer website, there are ~140,000 registered referees in the United States. As of 2021, the primary way for individuals to enter PRO2 is via recommendation from someone in their state and then being seen and approved by a member of the PRO2 committee. These officials are then given assignments in the USL-C, USL-L1 or NWSL. PRO2 coaches will sometimes watch some of their matches and provide feedback.

The committee then meets at the end of the year to determine if officials should be promoted or demoted. This process consists of the committee going through each person name by name and discussing their season. Historically, it had typically been done with very little reliable data. The process of coaching and improving the officials is primarily undertaken by the two PRO2 coaches. They have a handful of officials they work closely with when setting goals for the season and striving to improve the officials' performances week to week. One thing that has been missing for this process is a detailed understanding of where an official's strengths and weaknesses lie and how that is compared to officials within their tier, as well as t officials above or below them. Currently there is no record whether an official is improving or regressing season to season. This project was developed as an attempt to provide a data-driven way to make promotion/relegation decisions as well as improve on the individualization of coaching by providing an understanding of each official's strengths and weaknesses.
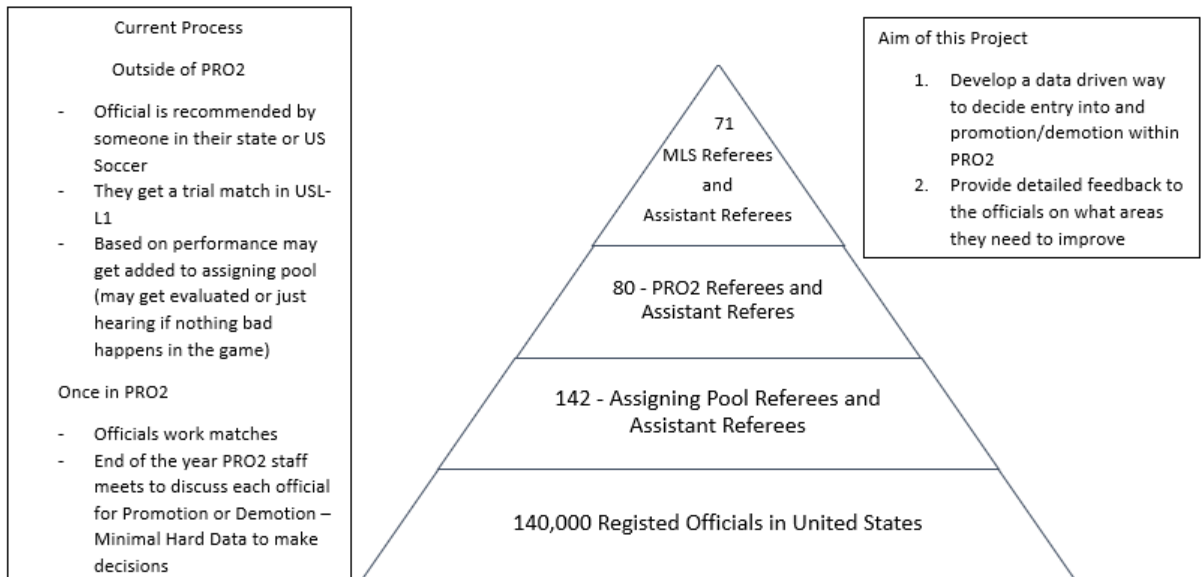
Figure 5 - Breakdown of the Number of Referees of Different Categories in the US and the process for moving from outside to inside the PRO2 group and then moving from PRO2 to MLS. (2021)

The aims of this project were 1. Conduct study 1 to determine what the definition of a successful referee performance is and what are the characteristics of a successful referee from a broad range of stakeholders in Major League Soccer (MLS). 2. Use the results of study 1 in collaboration with PRO2 coaches to develop a way to assess the referees: a) to differentiate between different levels of match official and b) to determine areas of strength and weakness for each match official. 3. Develop a battery of off-field methods to assess determinants of referee performance, based on items deemed important during the Delphi study which: a) have a strong correlation with the in-game assessment and b) differentiate between distinct levels of official.

The Problem
- Given the numbers in Figure 5 how do we know:
- 1. Who to promote inside of PRO2
- 2. Who do we invite into PRO2

Solution Part 1
- Define what we are looking for in a match offical.
- What are the components of a succesful referee.

Solution Part 2b
- Assessing these components off of the field

Solution Part 2a
- Designing an In-Match Assessment process
- Assessing these components during a match to differentiate between tiers

Figure 6 – Project Flow Chart displaying how each study is connected and progresses through each stage of the project.

6.3 The Solution

There were some aspects of this project which were successful, aim 1, aim 2a and aim 2b were met during this project. There were also aspects which were unsuccessful, aim 3a and aim 3b were not met. The following gives an overview of each project and the approach used to achieve each aim.

6.3.1 Part 1 – Chapter 3

The first step in this project was to develop an understanding of what it meant for a referee to be successful as well as what were the characteristics of a successful referee. This was accomplished utilizing the Delphi method to collect the perspectives of MLS Coaches & General Managers, MLS League Officers, PRO Referees, PRO Assistant Referees, PRO Assessors and PRO Staff members. There were 38 respondents in part 1 and 52 in part 2 including former National Team coaches and players as well as former and current FIFA officials. The definition of a successful referee performance was that they communicate with the players and coaches, no one talks about their impact on the game/they don't negatively impact the game, they make very few errors/all major decisions are correct, they protect the health and safety of the players and both teams feel like they received equal treatment. The 26 characteristics of a successful referee can be seen above in Table 3.

This project achieved aim 1. The definition of a successful referee performance was developed, and the components of a successful referee were developed. The key takeaways are a definition for what it means for an official to be successful as well as a multi-dimensional model of what characteristics are important for an official to develop to be successful in the MLS. This leads into the second project.

6.3.2 Part 2a Chapter 4

The development of a method for the on-field assessment of officials was based on this multi-dimensional model of a successful referee. The model was developed in collaboration with the PRO2 staff using the results of the Delphi study in combination with their domain expertise. The end product was the NOVA system. The items of the NOVA and their counterparts for the Delphi study can be seen in Table 6. The coaches practiced using the NOVA on matches from the previous season and developed a strong ICC. They then implemented it for the 2021 USL-C Season. The three coaches evaluated the performance of 55 officials over 304 matches with the number of matches evaluated per official ranging from 1 to 12 matches. The average and median number of matches evaluated per official was 5. The number of matches evaluated by each coach was 58, 116, 130, respectively.

The NOVA assessment tool was able to differentiate between level of official with tier A & B scoring higher than tier C, tier D and tier P for total game score. There were differences between level of official for 10 out of 18 items. This allows the coaches to understand where each official sits relative to the other officials for each item. In addition, it shows where trends may be for what causes the differences between levels of official.

This project achieved aim 2a and 2b. The assessment tool developed was able to differentiate between level of official and was able to display the individual strengths and weaknesses of each official. The key take-away was that the NOVA assessment tool was able to differentiate between tiers of official. However, to watch each match in full, score it and debrief with the crew from each match could detract from the ability to coach during the season. The next step is to try to develop an off-field method for evaluating the referees that compares favourably to the NOVA scores.

6.3.3 Part 2b Chapter 5

A battery of tests was developed to examine each item in the psychological and professional categories, knowledge of laws of the game, endurance ability and communication and teamwork. The psychological characteristics were evaluated using a combination of six tests the TOPS 2-S (composure and focus), Vealey's Trait Confidence Questionnaire (confidence), MTQ-18 (resilience), BIA/BIS (courage), Measure of Transformational Leadership (leadership), and the Big 5 Personality Questionnaire (strong personality). They were modified to remove repeated or irrelevant questions. There were individual questions developed for each of the items in the professional category. A test was developed with the help of the PRO Head of Coaching, Education and Evaluation to evaluate knowledge of laws of the game. An official's ability to utilize communication and teamwork was evaluated using peer scoring and endurance ability was evaluated using a 4-minute submaximal assessment as well as a 2000m time trial.

The off-field assessment was not as effective as the on-field assessment. There was no difference between tiers of official for any of the psychological or professional components. There was no difference between tiers for knowledge of laws of the game. There was a difference between tier B and D for communication and teamwork, but a larger sample is needed to have more confidence in the results. There was no difference between tiers for the 2000m time trial. There was a difference between tiers for the 4-minute submaximal assessment. It is possible that the difference between the MLS officials and tier A, B, and C was due to experience with the test as their heart rate was higher after the 4-minute run and 1-minute standing recovery, but the size of the heart rate recovery was the same. Tier D however also had a lower heart rate recovery, so it is possible that there is a difference between endurance ability between the MLS officials and tier D officials.

The comparisons between NOVA and the assessment were moderate for the knowledge of laws of the game and low for communication and teamwork.

This study did not achieve aim 3a or 3b. There was not a strong correlation between the off-field assessment and the NOVA Score nor were the off-field assessments able to differentiate between level of official. It is possible that the skills of a match official are only displayed on the field of play and cannot be broken into separate categories and evaluated off the field.

Currently it is recommended that only the NOVA score can be added to the promotion/demotion discussion process.

6.4 Application

The NOVA Scores were used at the end of 2021 season as pieces to the puzzle during discussions for promotion to the MLS level and for placement into the tiers within PRO2. The officials were grouped based on a combination highest total score, highest average score and lowest total score. There were a few officials who were promoted up to a higher tier based on their combination of scores being above the others in their current tier, and one official was demoted based on their combination of scores being below their current tier. There was an official removed from the assigning pool based on having the lowest combination of scores for the season. The NOVA score is not the only factor considered as there are other off-field factors which can contribute as well as trajectory of performance over multiple seasons.

Each referee in tiers A & B who officiated a match in the USL-C was given a report that detailed where their strengths and weaknesses were relative to officials in each tier. The coaches can use this to guide their professional development plan for the next season.

The mission of PRO2 is to find and develop officials for promotion into PRO while also supplying officials for USL and NWSL matches. This project allows them to individualize the coaching they provide to each official. Hopefully this will increase the rate at which they can develop officials for the MLS level as well as enhance the quality of officiating services provided to USL and NWSL matches.

6.5 Reflection

The process of the Professional Doctorate has been more difficult mentally than I had anticipated. It has involved several different emotions throughout this journey. I will admit I am not very good at written reflective practice. I spend a lot of time reflecting in my head, but I do not record my thoughts very well. I have tried a number of different methods of reflection, but as of now none have proven effective. I am much better at writing out prospective plans as opposed to writing retrospectively. The majority of my focus will be on the emotions felt during different aspects of this process, there are a number of these that I

can vividly remember. I will talk about the areas where I think I improved as well as the areas where I think I performed below my expectations

The best place to start is the beginning. I decided to embark on this project because I felt frustrated, at not being able to understand what targets we were aiming at for our officials. I could work to improve an official's fitness but how did I know that it was contributing to their performance on the field. I felt frustrated with the promotion/demotion decisions because I felt the decisions were not as data driven as they could be. I channelled this frustration into trying to answer the question of what does it mean for a referee to be successful? This led me to starting the program at Liverpool John Moores University.

 The next emotion is the excitement I felt when this project initially started to come together. I can vividly remember the excitement as I flew to England to register for the program and meet my advisors. I remember staying up through the night writing the first draft of my introduction and methods section. Looking out the big windows of my Airbnb overlooking the city the excitement of starting out on a new adventure keeping me awake. The feeling of excitement while working on the introduction and methods for the first project very quickly turned to feeling that I didn't belong, and I had made a terrible mistake when I saw all the red marking the copy sent back to me by my advisors. I went through a number of emotions before I could even read the edits. I felt like quitting the program, I felt angry, I felt inadequate, and I felt embarrassed. I very quickly closed the document, and it took a few days before I could open it. Once I began working through the edits one by one, I started to feel better. I realized that it wasn't as bad as I had thought, and it was mostly a matter of needing to shift my style from conversational to scientific (an area in which I still struggle). These feelings would still come back each time I received edits from my advisors but slowly over time they decreased in severity as I grew to understand that this is part of the process and not an attack on me or a reflection of who I am. It is part of the process of improving and getting better, and also a key part of the scientific process. You cannot just throw something together in a couple days. It takes weeks and weeks of edits and feedback until it gets melded into a finished product.

The next rush of excitement came when I received the first round of responses to project 1. It was a realization that it was actually happening and seeing the wide range of respondents

made me excited that others were viewing it as worth their time. This was again followed by a period of questioning and frustration as I felt things were moving slower than I would like. I struggled with the slower pace of academia. It was difficult for me slow down and go through each step to complete things on the academic side when I really wanted to just move from one step to the next as quickly as possible. I understand now the importance of slowing down and refining everything prior to moving on to the next step.

The third rush of excitement came when I was looking at the NOVA Scores grouped by tier. I saw that there were differences between the tiers for 10 of the items. I thought that it might have actually worked and all of the effort over the previous year had been worth it. This excitement was again followed by strong negative emotions was after reading through other prof doc papers to get an idea for formatting and organization. I again felt inadequate and embarrassed. The papers I was reading were substantially longer and more advanced than my project. I have since been assured that it is not the case, but it took a while for those feelings to subside. I am sure I will feel excitement at least one more time when I hand in the completed version of this document.

The actual work of the program was not particularly difficult, but the mental and emotional challenges greater than I had anticipated and took a lot to overcome. Another area with which I struggled initially was the difference in communications styles. I am used to much more direct feedback such as, "this is bad", "this needs to change", "this needs to go", etc. It took time to get used to "this is good" hiding a number of changes that need to be made, or "this is fine" meaning this needs to go or be completely reworked. It has taken time to get used to, but I have grown to understand that style of communication much more through this process.

I learned that I work more effectively from a writing standpoint when I work in focused bursts as opposed to trying to work little by little. It takes me a long time to regain my place when I try to do only a little at a time. The information flows much more coherently and effectively when I work in a focused burst. I found that using artificial deadlines of when I would get a draft to my advisors was effective around 70% of the time and was effective for instigating the burst. However, I did find when I missed one of those deadlines, my

motivation decreased. I needed to be careful that my deadlines were achievable, otherwise I would be more delayed in sending in a draft.

One skill that I feel I was able to develop during this project was to determine a problem and design a methodological approach to solve that problem via experimentation. I also learned the value in slowing down to make sure you get things correct as opposed to simply trying to get it done as quickly as possible.

While my thinking has improved, I still have a long way to go in effectively communicating my thoughts in a clear and concise manner. My writing still has a long way to go. I need to continue to focus on writing in a less conversational manner. I feel that this has been a very rewarding journey. I will continue to use this experience to help me grow in my career as a research practitioner.

6.6 NOVA Feedback

While the NOVA system looks good from a research perspective it is important to understand how those who used it and those who were assessed by it felt about it. A survey was sent out to all officials who were assessed via the NOVA system and the coaches who used it were asked to provide feedback. They were asked to rank NOVA 1 Very Detrimental - 5 Very Beneficial. They were then asked to provide any positive or negative written feedback, what changes they think should be made and if it were feasible to utilize for the 2022 season would they like to see it utilized. 17 Referees provided feedback on the NOVA system; 7 rated it Beneficial, 5 rated it Very Beneficial and 5 rated it Neither Beneficial nor Detrimental. Two coaches rated it Very Beneficial and 1 rated it Beneficial.

Common areas of positive feedback were ability to see individual strengths and areas for development, very detailed, and creation of a merit list by category. Common critiques were double penalized in some categories, tough to know why a given score was received, not all games were evaluated, too many items, more categories needed and confirmation bias as those who tiered the officials also assessed them. Recommended changes were inclusion of a comments section, more criteria, less criteria, more emphasis on positioning, place to see all scores and video clips to gain insight into why you received the score you did, more clarity, engagement of the full crew and referees also score themselves. 13 of 20 respondents responded that they would like to see it utilized in 2022, 3 responded yes, with

changes and all matches assessed, 3 did not answer the question and 1 did not give a yes or no answer.

Overall, the feedback from the officials and coaches was positive. A couple of the recommended changes could provide ways to improve NOVA from a teaching/development perspective.

6.7 Future Research

There are four pathways of future research to build off this project. 1. Repeat the process with assistant referees. 2. Refine and improve the NOVA system with another season of data. 3. Continue to try to find off-field methods for differentiating between tiers of official. 4. Track the results of current education methods based on the different items in the NOVA system and determine if the current education methods are effective.

Assistant referees have different on-field technical and physical requirements compared to referees, so it is important to understand what it means for an assistant referee to be successful and not assume that it is the same as for a referee. Just as with referees we need a data-informed way to decide which assistant referees should be promoted/demoted.
It is important to see if the results hold up in a second season of matches and reduce the likelihood of it being a fluke. Attempting to use a more even sample of matches where the range for number of matches observed is lower than 12 to 1 could be beneficial.
The time-cost of having to watch and score matches for anyone being considered for inclusion into PRO2 is cost prohibitive given the current ration of coaches to officials. to rely on. Methods to screen candidates need to be developed. While the current attempt was unsuccessful it is important to keep trying to develop a realistic approach.
PRO invests significant time into providing education for the match officials. Currently there is no way to track if the current educational methods are effective. Utilizing the NOVA to track the different items associated with the education sessions could allow for evaluation of their effectiveness.

6.8 Conclusion

The aims of this project were 1. Conduct study 1 to determine what the definition of a successful referee performance is and what are the characteristics of a successful referee from

a broad range of stakeholders in Major League Soccer (MLS). 2. Use the results of study 1 in collaboration with PRO2 coaches to develop a way to assess the referees: a) to differentiate between different levels of match official and b) to determine areas of strength and weakness for each match official. 3. Develop a battery of off-field methods to assess determinants of referee performance, based on items deemed important during the Delphi study which: a) have a strong correlation with the in-game assessment and b) differentiate between distinct levels of official.

There were some aspects of this project which were successful, aim 1, aim 2a and aim 2b were met during this project. There were also aspects which were unsuccessful. Aim 3a and aim 3b were not met.

The successful aspects of this project provide a strong foundation to continue to improve the quality of officiating in professional soccer in the United States. It has provided a data-driven way for the official to be evaluated. It has helped to standardize the decision-making process for promotion/relegation as there are data for each decision that is made. This project hopefully moves referee assessment and development forward allowing it to keep up with the growing demands of the game worldwide.

Chapter 8

References

References

Armatas, V. and Pollard, R. (2014) 'Home advantage in Greek football', *European Journal of Sport Science*, 14(2), pp. 116–122. doi:10.1080/17461391.2012.736537.

Armenteros, M. *et al.* (2018) 'The training of soccer assistant referees beyond on-field experience: the use of the Interactive Video Test', *International Journal of Computer Science in Sport*, 17(2), pp. 163–174. doi:10.2478/ijcss-2018-0009.

Bangsbo, J., Nørregaard, L. and Thorsø, F. (1991) 'Activity profile of competition soccer', *Canadian journal of sport sciences = Journal Canadien des sciences du sport*, 16, pp. 110–116.

Baptista, A.M.G. *et al.* (2021) 'Association between Clinical Vision Measures and Visual Perception and Soccer Referees' On-field Performance', *Optometry and Vision Science*, 98(7), pp. 789–801. doi:10.1097/OPX.0000000000001722.

Barbero-Álvarez, J. *et al.* (2012) 'Physical and Physiological Demands of Field and Assistant Soccer Referees During America's Cup', *The Journal of Strength & Conditioning Research*, 26(5), pp. 1383–1388. doi:10.1519/JSC.0b013e31825183c5.

Betsch, C. (2004) 'Präferenz für Intuition und Deliberation (PID)', *Zeitschrift für Differentielle und Diagnostische Psychologie*, 25(4), pp. 179–197. doi:10.1024/0170-1789.25.4.179.

Bizzini, M *et al.* (2009) 'Female soccer referees selected for the FIFA Women's World Cup 2007: survey of injuries and musculoskeletal problems', *British Journal of Sports Medicine*, 43(12), p. 936. doi:10.1136/bjsm.2008.051318.

Bizzini, M. *et al.* (2009) 'Injuries and musculoskeletal complaints in referees and assistant referees selected for the 2006 FIFA World Cup: retrospective and prospective survey', *British Journal of Sports Medicine*, 43(7), pp. 490–497. doi:10.1136/bjsm.2008.048314.

Bizzini, M. *et al.* (2011) 'Injuries of football referees: a representative survey of Swiss referees officiating at all levels of play', *Scandinavian Journal of Medicine & Science in Sports*, 21(1), pp. 42–47. doi:10.1111/j.1600-0838.2009.01003. x.

Bouzas-Rico, S. *et al.* (2021) 'Field-based tests for assessing fitness in referees: A systematic review', *Research in Sports Medicine (Print)*, pp. 1–19. doi:10.1080/15438627.2021.1895787.

Boyer, S., Rix-Lièvre, G. and Récopé, M. (2015) 'L'arbitrage de haut niveau, une affaire d'équipe', *Movement & Sport Sciences - Science & Motricité*, (87), pp. 91–101. doi:10.1051/sm/2014014.

Casajus, J.A. and Castagna, C. (2007) 'Aerobic fitness and field test performance in elite Spanish soccer referees of different ages', *Journal of Science and Medicine in Sport*, 10(6), pp. 382–389. doi: 10.1016/j.jsams.2006.08.004.

Castagna, C., Abt, G. and D'Ottavio, S. (2002) 'Relation between fitness tests and match performance in elite Italian soccer referees', *Journal of Strength and Conditioning Research*, 16(2), pp. 231–235.

Castagna, C., Abt, G. and D'Ottavio, S. (2004) 'Activity profile of international-level soccer referees during competitive matches', *Journal of Strength and Conditioning Research*, 18(3), pp. 486–490. doi:10.1519/1533-4287(2004)18<486: APOISR>2.0.CO;2.

Castillo, D. *et al.* (2019) 'Analysis of the success in soccer match officials' decisions during an international tournament according to contextual factors. [Análisis del éxito en las decisiones de los árbitros en partidos oficiales de fútbol durante un torneo internacional atendiendo a factores contextuales].', *RICYDE. Revista Internacional de Ciencias del Deporte. doi:10.5232/ricyde*, 15(57), pp. 225–234.

Catteeuw, P. *et al.* (2009) 'Decision-making skills, role specificity, and deliberate practice in association football refereeing', *Journal of Sports Sciences*, 27(11), pp. 1125–1136. doi:10.1080/02640410903079179.

Catterall, C. *et al.* (1993) 'Analysis of the work rates and heart rates of association football referees.', *British Journal of Sports Medicine*, 27(3), pp. 193–196.

Choi, H. and Chiu, W. (2017) 'Influence of the perceived organizational support, job satisfaction, and career commitment on football referees' turnover intention', *Journal of Physical Education and Sport*, 17, pp. 955–959. doi:10.7752/jpes. 2017.s3146.

Costa, E.C. *et al.* (2013) 'Monitoring External and Internal Loads of Brazilian Soccer Referees During Official Matches', *Journal of Sports Science & Medicine*, 12(3), pp. 559–564.

Dawson, P., Webb, T. and Downward, P. (2021) 'Abuse is not a zero-sum game! The case for zero tolerance of match official physical and verbal abuse', *European Journal of Sport Science*, pp. 1–8. doi:10.1080/17461391.2021.1881619.

De Oliveira, M.C, Orbetelli, R. and De Baros Neto, T.L. (2011) 'Call Accuracy and Distance from the Play: A Study with Brazilian Soccer Referees', *International Journal of Exercise Science*, 4(1), pp. 30–38.

Dohmen, T. and Sauermann, J. (2016) 'Referee Bias', *Journal of Economic Surveys*, 30(4), pp. 679–695. doi:10.1111/joes.12106.

D'Ottavio, S. and Castagna, C. (2001) 'Physiological load imposed on elite soccer referees during actual match play', Journal of Sports Medicine and Physical Fitness p.27-32.

Downward, P. and Jones, M. (2007) 'Effects of crowd size on referee decisions: analysis of the FA Cup', *Journal of Sports Sciences*, 25(14), pp. 1541–1545. doi:10.1080/02640410701275193.

Ekmekçi, Y.A.D., Kundakcı, N. and Ekmekci, R. (2020) 'Performance Evaluation of Basketball Referees with an Integrated MCDM Approach', Sport Mont, 18(2), pp. 95–102.

Fischer, K. and Haucap, J. (2021) 'Does Crowd Support Drive the Home Advantage in Professional Football? Evidence from German Ghost Games during the COVID-19 Pandemic', *Journal of Sports Economics*, p. 15270025211026552. doi:10.1177/15270025211026552.

Fowler, P.M. *et al.* (2017) 'Greater Effect of East versus West Travel on Jet Lag, Sleep, and Team Sport Performance', *Medicine & Science in Sports & Exercise*, 49(12), pp. 2548–2561. Available at: https://doi.org/10.1249/MSS.0000000000001374.

Garicano, L., Palacios-Huerta, I. and Prendergast, C. (2005) 'Favouritism under Social Pressure', *The Review of Economics and Statistics*, 87(2), pp. 208–216.

Ghasemi, A. *et al.* (2011) 'Visual Skills Involved in Decision Making by Expert Referees', *Perceptual and Motor Skills*, 112(1), pp. 161–171. doi: 10.2466/05.22.24.27.PMS.112.1.161-171.

Giel, T. and Breuer, C. (2020) 'The determinants of the intention to continue voluntary football refereeing', *Sport Management Review*, 23(2), pp. 242–255. doi: 10.1016/j.smr.2019.01.005.

Gilis, B. *et al.* (2006) 'Interpretation and application of the laws of the game in football incidents leading to player injuries', *International Journal of Sport Psychology*, 37(2–3), pp. 121–138.

Gorczynski, P. and Webb, T. (2021) 'Developing a mental health research agenda for football referees', *Soccer & Society*, 22(6), pp. 655–662. doi:10.1080/14660970.2021.1952695.

Goumas, C. (2014) 'Home advantage and referee bias in European football', *European Journal of Sport Science*, 14(sup1), pp. S243–S249. doi:10.1080/17461391.2012.686062.

Helsen, W. and Bultynck, J.-B. (2004a) 'Physical and perceptual-cognitive demands of top-class refereeing in association football', *Journal of Sports Sciences*, 22(2), pp. 179–189. doi:10.1080/02640410310001641502.

Helsen, W. and Bultynck, J.-B. (2004b) 'Physical and perceptual-cognitive demands of top-class refereeing in association football', *Journal of Sports Sciences*, 22(2), pp. 179–189. doi:10.1080/02640410310001641502.

Hill, D. (2010) 'A critical mass of corruption: Why some football leagues have more match-fixing than others', *International Journal of Sports Marketing and Sponsorship*, 11, pp. 221–235. doi:10.1108/IJSMS-11-03-2010-B005.

Hong, E., Jeong, Y. and Downward, P. (2019) 'Perceived organizational support, internal motivation, and work–family conflict among soccer referees', *Managing Sport and Leisure*, 24(1–3), pp. 141–154. doi:10.1080/23750472.2019.1593049.

van der Horst, N. *et al.* (2017) 'Return to play after hamstring injuries in football (soccer): a worldwide Delphi procedure regarding definition, medical criteria and decision-making', *British Journal of Sports Medicine*, 51(22), pp. 1583–1591. doi:10.1136/bjsports-2016-097206.

Hossner, E.-J. *et al.* (2019) 'The role of viewing distance and viewing angle on referees' decision-making performance during the FIFA World Cup 2014', *Journal of Sports Sciences*, 37(13), pp. 1481–1489. doi:10.1080/02640414.2019.1570898.

Janse van Rensburg, D.C. *et al.* (2021) 'Managing Travel Fatigue and Jet Lag in Athletes: A Review and Consensus Statement', *Sports Medicine*, 51(10), pp. 2029–2050. Available at: https://doi.org/10.1007/s40279-021-01502-0.

Jeffriess, M.D. et al. (2020) 'A preliminary, multidisciplinary investigation into the performance of professional referees in the Australian National Rugby League', p. 188.

Johnston, L. and McNaughton, L. (1994) 'The physiological requirements of Soccer refereeing', Australian Journal of Science and Medicine in Sport, 26(3–4), pp. 67–72.

Kittel, A. *et al.* (2019) 'Video-based testing in sporting officials: A systematic review', *Psychology of Sport and Exercise*, 43, pp. 261–270. doi: 10.1016/j.psychsport.2019.03.013.

Kleynen, M. *et al.* (2014) 'Using a Delphi Technique to Seek Consensus Regarding Definitions, Descriptions and Classification of Terms Related to Implicit and Explicit Forms of Motor Learning', *PLOS ONE*, 9(6), p. e100227. doi: 10.1371/journal.pone.0100227.

Kordi, R. *et al.* (2013) 'Incidence, Nature, and Pattern of Injuries to Referees in a Premier Football (Soccer) League', *Sports Health*, 5(5), pp. 438–441. doi:10.1177/1941738113481428.

Krustrup, P. *et al.* (2009) 'Activity profile and physical demands of football referees and assistant referees in international games', *Journal of Sports Sciences*, 27. doi:10.1080/02640410903220310.

Krustrup, P. and Bangsbo, J. (2001) 'Physiological demands of top-class soccer refereeing in relation to physical capacity: effect of intense intermittent exercise training', *Journal of Sports Sciences*, 19(11), pp. 881–891. doi:10.1080/026404101753113831.

Lago-Peñas, C. and Gómez-López, M. (2016) 'The Influence of Referee Bias on Extra Time in Elite Soccer Matches', *Perceptual and Motor Skills*, 122(2), pp. 666–677. doi:10.1177/0031512516633342.

Law 5 - The Referee | IFAB (no date). Available at: https://www.theifab.com/laws/latest/the-referee/#decisions-of-the-referee (Accessed: 7 March 2022).

Lex, H. *et al.* (2015) 'Influence of players' vocalisations on soccer referees' decisions', *European Journal of Sport Science*, 15(5), pp. 424–428. doi:10.1080/17461391.2014.962620.

Linstone, H. and Turoff, M. (1975) *The Delphi Method Techniques and Applications*. Available at: file:///C:/Users/jwestbrooks/Downloads/TheDelphiMethod%20(1).pdf.

Liu, T. *et al.* (2019) 'Exploring home advantage and quality of opposition interactions in the Chinese Football Super League', *International Journal of Performance Analysis in Sport*, 19(3), pp. 289–301. doi:10.1080/24748668.2019.1600907.

Loghmani, M., Cuskelly, G. and Webb, T. (2021) 'Examining the career dynamics of elite football referees: a unique identification profile', *Sport Management Review*, 24(3), pp. 517–542. doi:10.1080/14413523.2021.1879556.

MacMahon, C. *et al.* (2007) 'Decision-making skills and deliberate practice in elite association football referees', *Journal of Sports Sciences*, 25(1), pp. 65–78. doi:10.1080/02640410600718640.

Mallo, J. *et al.* (2007) 'Activity profile of top-class association football referees in relation to performance in selected physical tests', *Journal of Sports Sciences*, 25(7), pp. 805–813. doi:10.1080/02640410600778602.

Mallo, J. *et al.* (2009) 'Activity profile of top-class association football referees in relation to fitness-test performance and match standard', *Journal of Sports Sciences*, 27(1), pp. 9–17. doi:10.1080/02640410802298227.

Mallo, J. *et al.* (2012) 'Effect of positioning on the accuracy of decision making of association football top-class referees and assistant referees during competitive matches', *Journal of Sports Sciences*, 30(13), pp. 1437–1445. doi:10.1080/02640414.2012.711485.

Martínez, N., Botey Fullat, M. and Arce, S. (2021) 'Analysis of Burnout and Psychosocial Factors in Grassroot Football Referees', *International Journal of Environmental Research and Public Health*, 18, p. 1111. doi:10.3390/ijerph18031111.

Mascarenhas, D., Collins, D., and Mortimer (2005) 'The Art of Reason versus the Exactness of Science in Elite Refereeing: Comments on Plessner and Betsch (2001)', *Journal of Sport Exercise Psychology*, 24, pp. 328–333. doi:10.1123/jsep.24.3.328.

Mascarenhas, D.R. *et al.* (2009) 'Physical performance and decision making in association football referees: A naturalistic study.' Available at: http://collections.crest.ac.uk/314/ (Accessed: 1 October 2017).

Mather, G. and Breivik, S. (2020) 'Is the perception of intent by association football officials influenced by video playback speed?', *Royal Society Open Science*, 7(6), p. 192026. doi:10.1098/rsos.192026.

McCarrick, D. *et al.* (2021) 'Home advantage during the COVID-19 pandemic: Analyses of European football leagues', *Psychology of Sport and Exercise*, 56, p. 102013. doi: 10.1016/j.psychsport.2021.102013.

Meckel, Y., Balikin, K. and Eliakim, A. (2020) 'pre-and mid-season repeated sprint ability of soccer referees from the first and second divisions', *International Journal of Sports Science & Coaching*, 15(1), pp. 82–90. doi:10.1177/1747954119887301.

Mendes, S. *et al.* (2021) 'The Portuguese referee performance model', *International Journal of Sports Science & Coaching*, 16(5), pp. 1117–1125. doi:10.1177/17479541211006321.

Moreira, P. *et al.* (2020) 'Comparison of the gaze behaviour of soccer referees of different levels', *Journal of Physical Education and Sport*, 20, p. 1930. doi:10.7752/jpes.2020.04261.

Moriconi, M. (2020) 'Deconstructing match-fixing: a holistic framework for sport integrity policies', *Crime, Law and Social Change*, 74(1), pp. 1–12. doi:10.1007/s10611-020-09892-4.

Moriconi, M. and De Cima, C. (2021) 'Why some football referees engage in match-fixing? A sociological explanation of the influence of social structures', *International Journal of Sport Policy and Politics*, 13(4), pp. 545–563. doi:10.1080/19406940.2021.1928731.

Moriconi, M. and Teixeira-Diniz, R. (2016) 'Mitos, percepciones y actitudes frente a los resultados combinados en Portugal: un estudio con énfasis en los árbitros y los hinchas', *Mitos, percepciones y actitudes frente a los resultados combinados en Portugal: un estudio con énfasis en los árbitros y los hinchas*, (82), pp. 145–164. doi:10.7458/SPP2016827159.

Nevill, A.M. *et al.* (2017) 'Inconsistency of decision-making, the Achilles heel of referees', *Journal of Sports Sciences*, 35(22), pp. 2257–2261. doi:10.1080/02640414.2016.1265143.

Nevill, A.M., Balmer, N.J. and Mark Williams, A. (2002) 'The influence of crowd noise and experience upon refereeing decisions in football', *Psychology of Sport and Exercise*, 3(4), pp. 261–272. doi:10.1016/S1469-0292(01)00033-4.

Nurcahya, Y. *et al.* (2021) 'Improving the performance of football assistant referee through video-based training', *Journal of Physics: Conference Series*, 1806(1), p. 012212. doi:10.1088/1742-6596/1806/1/012212.

Paes, M., Fernandez Perez, R. and Silva, A. (2011) 'Injuries occurred with Football (Soccer) Referees During the Match, Training and Physical Test. .', INTERNATIONAL SPORTMED JOURNAL, 12, pp. 74–84.Pettersson-Lidbom, P. and Priks, M. (2010) 'Behaviour under social pressure: Empty Italian stadiums and referee bias', *Economics Letters*, 108(2), pp. 212–214. doi: 10.1016/j.econlet.2010.04.023.

Picazo-Tadeo, A.J., González-Gómez, F. and Guardiola, J. (2017) 'Does the crowd matter in refereeing decisions? Evidence from Spanish soccer', *International Journal of Sport and Exercise Psychology*, 15(5), pp. 447–459. doi:10.1080/1612197X.2015.1126852.

Pina, J.A. e *et al.* (2019) 'To be or not to be an excellent football referee: different experts' viewpoints', *Journal of Sports Sciences*, 37(6), pp. 692–700. doi:10.1080/02640414.2018.1522940.

Plessner, H. and Betsch, T. (2000) 'Sequential Effects in Important Referee Decisions: The Case of Penalties in Soccer', *Sonderforschungsbereich 504, University Mannheim & Sonderforschungsbereich 504, University of Mannheim, Sonderforschungsbereich 504 Publications*, 23. doi:10.1123/jsep.23.3.254.

R. Bloggers (2020) 'Money, market value and competition in modern football | R-bloggers', 7 February. Available at: https://www.r-bloggers.com/2020/02/money-market-value-and-competition-in-modern-football/ (Accessed: 6 April 2022).

Raab, M. *et al.* (2020) 'The referee's challenge: a threshold process model for decision making in sport games', *International Review of Sport and Exercise Psychology*, 0(0), pp. 1–21. doi:10.1080/1750984X.2020.1783696.

Reiter, B. (2018) *Astroball: The New Way to Win It All.* United States: Crown.

Riiser, A. *et al.* (2018) 'The Construct Validity of the CODA and Repeated Sprint Ability Tests in Football Referees', *International Journal of Sports Medicine*, 39(8), pp. 619–624. doi:10.1055/a-0577-4073.

Russell, S., Renshaw, I. and Davids, K. (2020) 'Sport arbitration as an emergent process in a complex system: Decision-making variability is a marker of expertise in national-level football referees', *Journal of Applied Sport Psychology*, 0(0), pp. 1–25. doi:10.1080/10413200.2020.1831651.

Samuel, R.D., Tenenbaum, G. and Galily, Y. (2020) 'An Integrated Conceptual Framework of Decision-Making in Soccer Refereeing', *International Journal of Sport and Exercise Psychology* [Preprint]. doi:10.1080/1612197X.2020.1766539.

Schwarz, W. (2011) 'Compensating tendencies in penalty kick decisions of referees in professional football: evidence from the German Bundesliga 1963-2006', *Journal of Sports Sciences*, 29(5), pp. 441–447. doi:10.1080/02640414.2010.538711.

Scoppa, V. (2021) 'Social pressure in the stadiums: Do agents change behaviour without crowd support?', *Journal of Economic Psychology*, 82. doi: 10.1016/j.joep.2020.102344.

Skulmoski, G.J., Hartman, F.T. and Krahn, J. (2007) 'The Delphi Method for Graduate Research', *Journal of Information Technology Education: Research*, 6(1), pp. 1–21.

Slack, L. *et al.* (2013) 'Factors Underpinning Football Officiating Excellence: Perceptions of English Premier League Referees', *Journal of Applied Sport Psychology*, 25. doi:10.1080/10413200.2012.726935.

Slater, M. (2019) *Two officials assaulted a week, minimal BAME representation and a five-a-side ban - why there are so few referees*, *The Athletic*. Available at: https://theathletic.com/1368327/2019/11/13/two-officials-are-assaulted-every-weekend-and-there-hasnt-been-a-bame-referee-in-the-top-four-divisions-in-11-years/ (Accessed: 12 October 2021).

Soriano, G. *et al.* (2019) 'El lado positivo del arbitraje: Percepción de apoyo organizacional, motivación y compromiso', *Revista de Psicologia del Deporte*, 28, pp. 41–49.

Sors, F. *et al.* (2020) 'The sound of silence in association football: Home advantage and referee bias decrease in matches played without spectators', *European Journal of Sport Science*, 0(0), pp. 1–9. doi:10.1080/17461391.2020.1845814.

Spitz, J. *et al.* (2016) 'Visual search behaviours of association football referees during assessment of foul play situations', *Cognitive Research: Principles and Implications*, 1, p. 12. doi:10.1186/s41235-016-0013-8.

Spitz, J. *et al.* (2017) 'Does slow motion impact on the perception of foul play in football?', *European Journal of Sport Science*, 17(6), pp. 748–756. doi:10.1080/17461391.2017.1304580.

Spitz, J. *et al.* (2018) 'The impact of video speed on the decision-making process of sports officials', *Cognitive Research: Principles and Implications*, 3(1), p. 16. doi:10.1186/s41235-018-0105-8.

Spitz, J. *et al.* (2021) 'Video assistant referees (VAR): The impact of technology on decision making in association football referees', *Journal of Sports Sciences*, 39(2), pp. 147–153. doi:10.1080/02640414.2020.1809163.

Sutter, M. and Kocher, M.G. (2004) 'Favouritism of agents – The case of referees' home bias', *Journal of Economic Psychology*, 25(4), pp. 461–469. doi:10.1016/S0167-4870(03)00013-8.

Unkelbach, C. and Memmert, D. (2008) 'Game Management, Context Effects, and Calibration: The Case of Yellow Cards in Soccer', *Journal of Sport and Exercise Psychology*, 30(1), pp. 95–109. doi:10.1123/jsep.30.1.95.

Unkelbach, C. and Memmert, D. (2010) 'Crowd Noise as a Cue in Referee Decisions Contributes to the Home Advantage', *Journal of Sport & Exercise Psychology*, 32(4), pp. 483–498.

Visschers, J., Paoli, L. and Deshpande, A. (2020) 'Match-fixing: Football referees' attitudes and experiences', *Crime, Law and Social Change*, 74(1), pp. 77–95. doi:10.1007/s10611-019-09880-3.

Webb, T. *et al.* (2020) 'An analysis of soccer referee experiences in France and the Netherlands: Abuse, conflict, and level of support', *Sport Management Review*, 23(1), pp. 52–65. doi: 10.1016/j.smr.2019.03.003.

Webb, T. (2021) 'The future of officiating: analysing the impact of COVID-19 on referees in world football', *Soccer & Society*, 22(1–2), pp. 12–18. doi:10.1080/14660970.2020.1768634.

Webb, T., O'Gorman, J. and Markham, L. (2021) 'Striving for excellence: talent identification and development in English football refereeing', *European Sport Management Quarterly*, 0(0), pp. 1–19. doi:10.1080/16184742.2020.1864748.

Weston, M. *et al.* (2006) 'The effect of match standard and referee experience on the objective and subjective match workload of English Premier League referees', *Journal of Science and Medicine in Sport*, 9(3), pp. 256–262. doi: 10.1016/j.jsams.2006.03.022.

Weston, M. *et al.* (2007) 'Analysis of physical match performance in English Premier League soccer referees with particular reference to first half and player work rates', *Journal of Science and Medicine in Sport*, 10(6), pp. 390–397. doi: 10.1016/j.jsams.2006.09.001.

Weston, M. *et al.* (2008) 'Analysis of physical match performance in English Premier League soccer referees with particular reference to first half and player work rates', *Journal of science and medicine in sport / Sports Medicine Australia*, 10, pp. 390–7. doi: 10.1016/j.jsams.2006.09.001.

Weston, M. *et al.* (2009) 'Relationships among field-test measures and physical match performance in elite-standard soccer referees', *Journal of Sports Sciences*, 27(11), pp. 1177–1184. doi:10.1080/02640410903110982.

Weston, M. *et al.* (2010) 'Ageing and physical match performance in English Premier League soccer referees', *Journal of Science and Medicine in Sport*, 13(1), pp. 96–100. doi: 10.1016/j.jsams.2008.07.009.

Weston, M. *et al.* (2011) 'Changes in a Top-Level Soccer Referee's Training, Match Activities, and Physiology Over an 8-Year Period: A Case Study', *International Journal of Sports Physiology and Performance*, 6(2), pp. 281–286. doi:10.1123/ijspp.6.2.281.

Weston, M. *et al.* (2012) 'Science and Medicine Applied to Soccer Refereeing', *Sports Medicine*, 42(7), pp. 615–631. doi:10.2165/11632360-000000000-0

Defining A Successful Referee Performance

Introduction

You are being contacted due to your role within Major League Soccer. This project is a partnership between Liverpool John Moores University and the Professional Referee Organization Performance and Analysis Department. The first aim of this project is to create a unified definition of what constitutes a successful referee performance. The second aim of this project is to create a list of key characteristics and abilities that make up a successful referee.

Background

Although it is clear a referee can have a major impact on the outcome of a football match, there is currently no clear definition of what makes a successful referee performance in the MLS. According to the International Football Association Board (IFAB), the current duties of a referee are to "1. Enforce the laws of the game. 2. Control the match in cooperation with the other match officials. 3. Act as timekeeper, keeps a

record of the match and provide the appropriate authorities with a match report including information on disciplinary action and any other incidents that occurred before, during or after the match. 4. Supervise and/or indicate the restart of play." In order to improve the quality of officiating there needs to be a definition what it means for a referee to be successful. The end goal is to utilize the results of this project to guide the training and selection of referees for Major League Soccer to enhance the profile of the league.

You are part of a panel consisting of MLS Team General Managers, MLS Coaches, MLS Players, MLS Officers, PRO Referees, PRO Management Staff, PRO Assessors, PRO Assistant Referees and Developmental Referees. As subject matter experts you are asked to participate by responding to three brief questionnaires.

Questionnaire 1. A. What is your definition of a successful referee performance?

B. What are the **characteristics and abilities** of a successful referee?

Questionnaire 2. How would you rank these answers

(1-5) Questionnaire 3. How would you rank these

answers (1-5)

After the third questionnaire's results are compiled, it is anticipated that the panel of experts' responses will form a consensus in answering Question 1, parts A and B.

This study has been ethically approved by Liverpool John Moores University Research Ethics Committee Reference number:19/SPS/053

I hereby agree to participate in this research project.
Yes
No

Please select your role in relation to Major League Soccer

If you were/are a referee/assistant referee, what was the highest level you officiated and how many years did you officiate?

If you were/are a coach, what is the highest level you have coached and for how many years did you coach?

If you were/are a player, what is the highest level you played and how many years did you play?

Questionnaire 1

Thank you for taking the time to participate in this research project. The first part of this project consists of answering two questions. In the appropriate box below please provide your answers to the two questions.

Q1 A. What is your definition of a successful referee performance?

Q1 B. What characteristics (e.g., mental toughness, composure, focus etc.) or abilities (e.g., law knowledge, positioning, sprinting speed etc.) make a referee successful?
Please list as many characteristics and abilities as you can.

Final Page

Thank you for taking the time to complete the 1st of 3 questionnaires. Questionnaire 2 will be sent to you in early January.

Defining a Successful Referee Performance Part 2

Page 1: Introduction

Thank you for taking the time to complete part 1 questions A and B. I have consolidated your answers into themes. For part 2 of the survey, the goal is to get consensus on which themes should be included in defining successful referee performance. A consensus will be formed by having each item ranked on a 1-7 scale with 1 = not important and 7= Extremely important.

Below is a reminder as to the purpose of this project.

This project is a partnership between Liverpool John Moores University and the Professional Referee Organization Performance and Analysis Department. The first aim of this project is to create a unified definition of what constitutes a successful referee performance. The second aim of this project is to create a list of key characteristics and abilities that make up a successful referee.

Background

Although it is clear a referee can have a major impact on the outcome of a football match, there is currently no clear definition of what makes a successful referee performance in the MLS. According to the International Football Association Board (IFAB), the current duties of a referee are to "1. Enforce the laws of the game. 2. Control the match in cooperation with the other match officials. 3. Act as timekeeper, keeps a record of the match and provide the appropriate authorities with a match report including information on disciplinary action and any other incidents that occurred before, during or after the match. 4. Supervise and/or indicate the restart of play."

In order to improve the quality of officiating there needs to be a definition of what it means for a referee to be successful. The end goal is to utilize the results of this project to guide the training and selection of referees for Major League Soccer in order to enhance the profile of the league.

You are part of a panel consisting of MLS Team General Managers, MLS Coaches, MLS Players, MLS Officers, PRO Referees, PRO Management Staff, PRO Assessors, PRO Assistant Referees and PRO2 Referees. As subject matter experts you are asked to participate by responding to three brief questionnaires.

Question A. What is your definition of a successful referee

performance? Question B. What are the characteristics

and abilities of a successful referee?

Did you complete the 1st Survey?
Yes
No

Please select your role in relation to Major League Soccer
MLS Team General Manager
MLS Coach
MLS Officer
MLS Player
PRO Assistant Referee
PRO Assessor
PRO2 Referee
PRO Staff
PRO Referee

Page 2: Question A

Below are the themes collected from the responses to Question A. What is your definition of a successful referee performance?

A theme was created when there were more than 2 responses relating to that theme. Please rate the importance of each item.

Please don't select more than 1 answer(s) per row.

	1 – Not at all important	2 – Low importance	3 – Slightly important	4 – Neutral	5 – Moderately important	6 – Very important	7 – Extremely important
They control the game	☐	☐	☐	☐	☐	☐	☐
They communicate with the players	☐	☐	☐	☐	☐	☐	☐
No one talks about their impact on the game/they don't negatively impact the game	☐	☐	☐	☐	☐	☐	☐
They correctly enforce the laws of the game	☐	☐	☐	☐	☐	☐	☐

They make very few errors/all major decisions are correct	☐	☐	☐	☐	☐	☐	☐
They protect the health and safety of the players	☐	☐	☐	☐	☐	☐	☐
Each team feels like they received equal treatment	☐	☐	☐	☐	☐	☐	☐

Page 3: Question B

Below are the themes collected from the responses to Question B. What are the characteristics and abilities of a successful referee?

The responses were divided into 4 major categories: Physical, Technical, Psychological, and Professional. The responses were then combined to create a theme. A theme was created when there were more than 2 responses relating to that theme.

Please rate the importance of each item.

Physical Characteristics and Abilities

Please don't select more than 1 answer(s) per row.

	1 – Not at all important	2 – Low importance	3 – Slightly important	4 – Neutral	5 – Moderately important	6 – Very important	7 – Extremely important
Physically Capable	☐	☐	☐	☐	☐	☐	☐
Correct Movement	☐	☐	☐	☐	☐	☐	☐
Correct Positioning	☐	☐	☐	☐	☐	☐	☐
Correct Use of Body Language	☐	☐	☐	☐	☐	☐	☐

Technical Characteristics and Abilities

Please don't select more than 1 answer(s) per row.

	1 – Not at all important	2 – Low importance	3 – Slightly important	4 – Neutral	5 – Moderately important	6 – Very important	7 – Extremely important
Strong Communicator	☐	☐	☐	☐	☐	☐	☐
Football Understanding	☐	☐	☐	☐	☐	☐	☐
Knowledge of the Laws of the Game	☐	☐	☐	☐	☐	☐	☐
Ability to Apply the Laws of the Game	☐	☐	☐	☐	☐	☐	☐
Ability to Anticipate Play	☐	☐	☐	☐	☐	☐	☐
Ability to Manage Players/Coaches	☐	☐	☐	☐	☐	☐	☐
Makes the Correct Decisions	☐	☐	☐	☐	☐	☐	☐
Utilizes Teamwork	☐	☐	☐	☐	☐	☐	☐
Has Experience	☐	☐	☐	☐	☐	☐	☐

Psychological Characteristics and Abilities

Please don't select more than 1 answer(s) per row.

	1 – Not at all important	2 – Low importance	3 – Slightly important	4 – Neutral	5 – Moderately important	6 – Very important	7 – Extremely important
Composed Under Pressure	☐	☐	☐	☐	☐	☐	☐
Focused	☐	☐	☐	☐	☐	☐	☐
Confident	☐	☐	☐	☐	☐	☐	☐
Resilient (Mentally Tough)	☐	☐	☐	☐	☐	☐	☐
Adaptable	☐	☐	☐	☐	☐	☐	☐
Courageous	☐	☐	☐	☐	☐	☐	☐
Ability to Lead	☐	☐	☐	☐	☐	☐	☐
Strong Personality	☐	☐	☐	☐	☐	☐	☐

Professional Characteristics and Abilities

Please don't select more than 1 answer(s) per row.

	1 – Not at all important	2 – Low importance	3 – Slightly important	4 – Neutral	5 – Moderately important	6 – Very important	E in
Trustworthy	☐	☐	☐	☐	☐	☐	☐
Committed	☐	☐	☐	☐	☐	☐	☐
Student Mentality	☐	☐	☐	☐	☐	☐	☐
Approachable	☐	☐	☐	☐	☐	☐	☐
Humble	☐	☐	☐	☐	☐	☐	☐
Mentally Prepared	☐	☐	☐	☐	☐	☐	☐

Page 4: Final page

Thank you for taking the time to complete Survey 2.