Evaluation of the physical fitness of referees in Iceland

A comparison between football and handball referees

Lokaverkefni í íþróttafráði BSc
Vor, 2018

Nafn nemanda: Jón Reynir Reynisson
Kennitala: 180892-2229

Nafn nemanda: Stefán Freyr Michaelsson
Kennitala: 050188-2809

Leiðbeinendur: Fannar Karvel Steindórsson og Kristján Halldórsson
Tækni- og verkfræðideild
ABSTRACT

Professional football referees need a standardized physical testing system that can be administered regularly to maintain quality assurance in game play decision making. A quantitative study for officiating upper league referee’s physical capabilities was performed during the school year of 2017/2018 at Reykjavik University in order to examine the physical capabilities of referees in Iceland’s top division football, all of whom are registered officials through the football association of Iceland, KSÍ (Knattspyrnusamband Íslands).

In order to achieve this, we tested football referees from KSÍ in two physical agility tests, T-test and the Yo-yo Intermittent Recovery level II test (Yo-yo II, Yo-yo IR II). We also compared the referees to semiprofessional football players, which the referees of Iceland would govern during matches. A comparison to other upper level referees in team handball, HSÍ (Handboltasamband Íslands), a sport popular in Europe, was also conducted to see if and what the difference in capabilities is between professions.

The results showed that both groups of referees were not as efficient as the players in all regards. The KSÍ referees did better than the HSÍ referees in the T-test with mean time of 10.07 ± 0.19 seconds for the KSÍ referees compared with 10.16 ± 0.15 seconds for the HSÍ referees. In contrast, the HSÍ referees did better in the Yo-yo II test, with mean distance of 420 ± 52 meters for the HSÍ referees compared with 364 ± 34 meters for the KSÍ referees.

These results show that the younger athletes are in better shape than the older referees which is in accordance with other studies. Furthermore, the study shows that although the KSÍ referees were measured during their off-season training they did better than the HSÍ referees in their mid-season training in one of the two tests. Therefore, it may be concluded that the KSÍ referees are in overall better shape than the HSÍ referees. This needs to be confirmed with a larger cohort and comparing timing as of beginning, mid-term and end of season for both groups.
## TABLE OF CONTENT

ABSTRACT .......................................................................................................................... 3

TABLE OF CONTENT ........................................................................................................ 4

LIST OF FIGURES ............................................................................................................... 5

LIST OF TABLES .................................................................................................................. 6

INTRODUCTION .................................................................................................................. 7

- The demands made on football referees ...................................................................... 8
- The physical ability of football referees .......................................................................... 12

THE OBJECTIVE OF THE STUDY ..................................................................................... 14

SUBJECTS AND METHODS ............................................................................................... 15

- Subjects .......................................................................................................................... 15

- Tests used in the study .................................................................................................... 16

  - Test variables ................................................................................................................ 16
  - The T-Test ....................................................................................................................... 17
  - The Yo-yo intermittent recovery test level II .............................................................. 18

- Statistical analysis ......................................................................................................... 20

RESULTS ............................................................................................................................ 21

- Comparison between KSÍ and HSÍ referees ................................................................. 21

  - T-test results .................................................................................................................. 21
  - Yo-yo II test results ....................................................................................................... 22

- Comparison between HSÍ referees and semi-professional handball players .............. 23

  - T-test results .................................................................................................................. 23
  - Yo-yo II test results ....................................................................................................... 24

- Comparison between KSÍ referees and semi-professional football players .............. 25

  - T-test results .................................................................................................................. 25

DISCUSSION ....................................................................................................................... 26

CONCLUSION ....................................................................................................................... 31

CITATIONS .......................................................................................................................... 32

APPENDIX .......................................................................................................................... 36
LIST OF FIGURES

Figure 1. T-test. ......................................................................................................................... 17
Figure 2. Speeds and difficulty levels of Yo-yo II test............................................................... 19
Figure 3. Yo-yo II Field Diagram. ............................................................................................ 19
Figure 4. T-test results for KSÍ and HSÍ referees. ................................................................. 21
LIST OF TABLES

Table 1. The study subjects

15
INTRODUCTION

The premise of the research being presented in this study was to establish a solidified system to evaluate top level referees in football. Football or soccer is widely regarded as the most played and most watched sport in the world. During the final game of the 2014 men’s World Cup in Rio de Janeiro, the German team beat Argentina 1-0 in overtime. The multitude of viewers numbered over 1 billion for people that tuned in to watch the game during its airing (FIFA.com, 2015).

The reason why football is so popular is most likely due to its simplicity; all anyone needs to play is a ball to kick and a simple set of markers for goal posts. Kids around the world play pickup games with limited amount of equipment and in that situation remains the sport’s success. Therefore, the sport at its most simple level has no boundaries based on cultural or financial situation. This implements equality across the board of social income for any one athlete who plays the game.

Even though the game is very simple to play it brings with it more in-depth rules as the caliber of the game increases. Football can be quite complex in the higher levels, which makes for a more developed and competitive settings. That is why skilled referees are needed to maintain a certain standard. This is secured by Fédération International de Football Association (FIFA) which has almost a million referees registered and able to oversee games of all levels and about 14,000 registered to officiate premier games (FIFA.com, e.d.).

With the number of viewers world-wide, the officiating association must provide employees that will ensure all the rules are adhered to, in order to maintain consistency in the equality of game play and quality of refereeing. This is accomplished by employing referees that have passed rigorous physical and literary testing which qualifies them to properly judge games in ascending calibers. The demand in importance and depth of knowledge of the full subject increases as a referee moves higher in ranks.
The demands made on football referees

When officiating a match, the referee should preferably be within 20 yards (18 meters) of the ball at any given moment. Therefore, the average distance covered by the main referee of a top league game is anywhere between 11 and 13 kilometers. On top of that, studies show that the referees are not simply jogging the duration of the match. On numerous occasions they reach speeds of over 20 km/h and one study found that they run at above 90% of max heart rate for over 600 meters each game, more than the average player on the field (Mallo, et al., 2009). In addition, the assistant referees run around 6 km on the side (Krustrup, 2009). Most referees can expect to at least run an average of 11 +/- 2.5 kilometers per game, that of which is 1.15 km are high speed running and 1.16 km is a side shuffle.

The demands of being a semi-professional and especially professional referee are very high. Everybody relies on the officials to ensure a game or match play is conducted in adherence to all rules and regulations that are set in place to ensure both teams are playing fairly in accordance to all the rules. As a referee you need to be able to make split second judgments of players and their conduct in game play. This can be a daunting task if a referee is not in adequate shape. The need for a referee to be in his best shape is one of the key factors of success when it comes to officiating a game. Obviously not all games are of the same caliber but the standards should remain the same. High quality games need high quality referees with increased physical capabilities to officiate a game more effectively (Mallo, et al., 2009). The cardiovascular capability of a referee needs to be relatively high for the game which lasts ninety minutes and can go into overtime. A referee must be able to maintain a clairvoyant state of mind that is un-interrupted due to lack of conditioning. In short, a referee needs to be in top shape. The higher the caliber of the game the higher the demands are that they can conduct a full match without personal interference due to lack of conditioning (Kalapotharakos, et al., 2006).

A referee’s job is to make calls on rules that are broken in game play and to conduct protocols that ensure fairness and equality in match play between teams. The demands can be very high. This depends on the individual referee, as well as all the stress that is input from external stimuli. A game can have a lot riding on its outcome.
For instance, one of the biggest games each year, in regards to money made, is the Championship play-off game that determines which team goes to play in the English Premier League next season. To begin with 42 teams play in the Championship League and the two highest at the end of the season are automatically promoted. The next four teams, those who finish the season in placements 3 to 6 then face off in a two-game series where the 3\textsuperscript{rd} place plays the 6\textsuperscript{th} and the 4\textsuperscript{th} place takes on the 5\textsuperscript{th}. The winner of these series then play each other in that most important game, the game that can win a club up to £200 million in television fees and other payments (Criddle, 2016).

If a referee is not in his/her best shape then they could make decisions that are possibly wrong or even unfair. This would make it that the referee is not capable to accept and analyze each team’s field presentation accordingly due to being out of shape. Resulting in the referee not being able to say they confidently know he/she did their best to officiate the game without any physical deterrents on their behalf. In addition, referees can experience other issues from the exterior stimulus that the environment of the games can present (Nevill, et al., 2002). A referee can experience anxiety from criticism from some of the athletes, coaches and even the crowd. While being able to reduce this criticism from the athletes and coaches with warnings if they fall under the scrutiny of what is considered unsportsmanlike. For those types of actions can be punished with proper discipline as far as the rules of the game go. However, the referee can hardly contain an angry crowd yelling at him. This presents us with an uncontrollable exterior stimulus that plays upon the referees ability to focus on his/her job requirements (Jones, et al., 2002).

The caliber of each game varies drastically with the time of the season and the setting of the game. In a setting where 70,000 fans are watching a game, the amount of pressure is very high. The athletes are presented with the drive and pressure to give their best effort in order to continue in their quest for the championships. The referees also feel the pressures of the setting for it is their judgments that can affect how the game will be played. For a referee to make quality decisions on whether or not to blow the whistle they must be in top physical form to be as close to the action as possible to ensure they make a quality decision. However, as a referee gets older and gains more experience, there is a decrease of kilometers run per game. The older, wiser referee
does not need to move as much as a young novice referee. This is most likely not just due to age, rather the experience of knowing when you need to be in a certain position and when it is not required, i.e. a greater understanding of the game (Casajus and Castagna, 2007).

A referee can be affected by a multitude of factors apart from their physical caliber. All exterior stimuli in their lives can be taken into consideration. A study conducted by Jens Bangsbo showed that there are a multitude of factors that can affect the decision making process of the referees (Bangsbo, et al., 2006). The theme of decision making that the referees must deal with falls into five major categories. These situations are ‘experience factors’, and crowd factors, player reaction, environmental factors, and crowd interaction represent a higher-order factor labeled ‘situational factors’ (Bangsbo et al., 2006; Lane, et al., 2006). All of these situations play a part in the decision making process. The experience factor can, for example, be described as how the referee feels in the games setting, which may be affected by whether they have been in the situation before and how their experience compares with previous games. If they have a previous experience, then the stress level is most likely to be less. The crowd factor, e.g. how are they reacting to the referees decisions, is an important factor. When thinking of football one of the first things that comes to mind is the massive fan base and how they are not always the nicest or most tranquil fans of the decisions of the referee (Nevill, et al., 2002). One catalytic decision from an official can set a whole crowd into a frenzy of joy or rage (Jones, et al., 2002).

There is a lot of pressure on each individual referee. Fortunately, there is not just one but four referees in each match. This gives them the opportunity to discuss with one another the decision that must be made (Mallo, et al., 2012). However, most likely, all referees, especially the main one, will experience some form of stress during match play. The main referee is the lead official in a football game. It is his/her decision of whether or not to make a call/judgment at any given moment during a match; thus, he bears the brunt of the responsibility. Nevertheless, his usually two or more assistants have their own responsibility during the match. While the main referee is in the middle of the action on the pitch, the two linesmen each guard one half of the court. Whilst everyone remembers a bad call from a referee, few take notice of how often they make
The right decision. A recent study of how many decisions a referee has to make during a match in the English Premier League revealed that on average the main referee has to make 245 decisions each game. The study also showed that 240 of those decisions were correct, giving the referee a score of 98%. The assistant referees were also taken into account, they made on average 50 decisions in each game and 49 of those were correct, also granting them a score of 98% (Gerrard, 2018).

The occupational demands for a referee are very high. The individual who is officiating the matches must possess knowledge of all the rules and regulations involving the game and when and when not to implement them. In addition, few other skill sets are also key factors to maintaining control of a game, such as strong verbal skills, leadership skills and of course decision making skills (Betsch and Haberstroh, 2005). These are very important characteristics to possess. Good leadership skills and communication skills can assist in the calming of situations where the opposing teams might be upset with one another so the players can avoid conflict. The decision making skills come into play at this point for if the athletes persist to have disorderly conduct then a warning must be issued. The referees must have a keen eye for details. An act of misconduct or a foul can be minimal, but a foul is still a foul, so they must keep a sharp sight on the players so they can provide fair game play. They must also assess misconducts that occur within the match play. In order to deliver an appropriate call, the referee must preferably be in the middle of the action. If a referee is not in the midst of the action his point of view might be skewed (Betsch and Haberstroh, 2005). This is why it is so important for a referee to have a clear mindset, good conditioning and preferably a good vision. As a rule, few players and coaches are going to argue with a referee that has great positioning and is close to the action even if he makes wrong decisions. However, as soon as the referee is out of position and making bad calls, he becomes questionable and will gain all sorts of criticism from those who are affected by his decisions (Krustrup, et al., 2002). Another key factor to be a good referee is working with a good set of assistant referees who are located on the side lines.

The duties of a referee span further than just following the rule book. Off the field the referee must avoid speaking with the media, which can be hard to avoid, especially if it is a popular game that has a large following. Also, they must be able to
keep a good track of time for some games might roll into an over time so they must decide how long a match may continue.

This summary of the demands made to referees during game plays show that there is a plethora of rules, regulations and standards that must be adhered and performed. The job of a referee is nothing to take lightly.

The physical ability of football referees

Part of the physical endurance is the ability to run long distances as the referees have to be able to move with the ball back and forth on the football field. A recent study found that the average referee in the Champions League covered 11.6 km per game which is on par with the players they are officiating (Weston, 2007). When age is taken into consideration it is quite impressive since the average referee at that level is almost 15 years older than the average player on the field (Casajus and Castagna, 2007).

There are many studies that have been done for endurance testing and many of these use standardized running tests, such as the 12 minute run test, 6x40 meter sprint test along with the 150 meter interval test (Weston, et al., 2009). These tests do provide information about the cardiovascular capabilities of the athletes who participate in them (Krstrup, et al., 2006). However, these tests are not designed to emphasize on the agility of the athlete. Of course each test result could be further developed to use the measurement for more than just distance and obtain other parameters, such as agility and anerobics. This might give information which could be used to suggest additional testing to further the development of the referee testing curriculum.

There are new and exciting tests being done in this field, including the T-test and Yo-yo tests, using referees as subjects (Bangsbo, et al., 1991). In many ways, professional referees can be viewed as athletes, and, therefore, they must have the same set of capabilities as the athletes they officiate, at least comparable and even better. Athletes and referees in top condition have been categorized with a minimum of 3 major abilities by a study conducted by Bangsbo et al. that describes what those are and how they apply to the development of an athlete (Bangsbo, et al., 2006). These tests includes the ability to perform a prolonged exercise, capability of high intensity sprinting, and the ability to develop a high power single direction output such as
kicking a ball or turning to sprint (Bangsbo, et al., 2006). These types of demands show the need for all styles of testing when it comes to evaluating and testing referees. The agility and endurance testing from a Yo-yo test and t-test would provide us with a greater understanding and measurable value of each referees capabilities.

For about the last twenty years there has been plenty of data collected about the conditioning of referees. These tests, however, have been very focused on the running capability of the referees. The tests have been a pass/ fail 150 meters run along with a repeated sprint ability (RSA) test. These are tests that emphasize on the cardiovascular capability of the referee that partakes in the test. As the difference between an athlete’s and a referee’s conditions vary greatly, so they should preferably be evaluated using different tests. As there is often a 15 years difference between referees and top-level athletes (Casajus and Castagna, 2007), a comparison of physical fitness between two groups of such great age difference might be difficult and not cast a light on the abilities of the older group. Therefore, in order to increase the understanding of capabilities of football referees, referees in a similar sport, with regards to physical demands and complexity were evaluated. The officiating association that was most eligible for the study was the Handball Association of Iceland (HSÍ, Handboltasamband Íslands). Handball is a similar sport in a lot of regards, fast paced, running based and two teams with a ball and goal objective.
THE OBJECTIVE OF THE STUDY

In light of the demands made on physical and physiological efficiency of football referees, and how little this has been studied here in Iceland. This study was set up with the research question to determine whether the present criteria for evaluating referees is adequate and whether it can be improved to increase the quality of match officiation.

The objective of the study was to develop a better understanding and further the standards at which professional and semi-professional football referees are measured. This might lead to standardization of the demands for what agility and aerobic endurance should be for a referee to officiate a game.
SUBJECTS AND METHODS

Subjects

The study group in this study included eighteen KSÍ referees (table 1), at the mean age of 27.5 years (ranging from 19 to 42). Within the league of referees are different levels or ranks of referees. The levels are 1-9, one being a FIFA referee and nine a youth recreational referee. We analyzed the soccer referees between levels 3 and 7 (Dómaralisti - Knattspyrnusamband Íslands“, 2018). As for the referees in Europe, they perform at about 14 different levels, with level 1 being the Premiere league and 2-4 being the English Football league. This digresses in scale and caliber the farther along the list you go. The bottom being non-officiating officials whom are considered qualified but non-practicing. For comparison, a control group of nineteen HSÍ referees were included, as both are involved in semi-professional leagues. As there is often a 15 years difference between referees and top level athletes, a comparison of physical fitness between the two groups might not wield the necessary emphasis on the abilities of the older referees. Therefore, in order to increase the understanding of capabilities of football referees, referees in a similar sport, with regards to physical demands and complexity were evaluated. Handball is a similar sport in a lot of regards, fast paced, running based and two teams with a ball and goal objective. Hence the decision to use handball referees as the chosen comparison group.

Table 1. The study subjects

<table>
<thead>
<tr>
<th>Subjects (n)</th>
<th>KSÍ referees</th>
<th>HSÍ referees</th>
<th>Football players</th>
<th>Handball players</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Males</td>
<td>16</td>
<td>19</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

In addition, two extra control groups were included, which were individuals from semi-professional teams, thirteen 1st division semi-professional football players.
who play for the team Þróttur Reykjavík and thirteen 2nd division semi-professional handball players representing the team Handboltafélag Kópavogs (HK).

**Tests used in the study**

The tests used in this study were chosen to evaluate aerobic, anaerobic and agility capabilities of all the groups.

To measure the athletes two tests were chosen, which will help determine the groups level of athleticism with numerical values represented in time spent completing or participating in a test. One of the tests chosen was the T-test, which is used to determine agility and speed. The T-test is a valid way to measure leg strength and agility (Pauole, et al., 2000). Another test was the Yo-yo-II test, which is used to determine cardiovascular endurance and anaerobic capability (Castagna, et al., 2006). The Yo-yo test was chosen as it can provide an accurate assessment that does not involve having to perform a lot of laboratory test that would provide very similar results.

These two tests were also chosen because they simple to conduct and administer and were practical to use when evaluating a multitude of athletes accurately at the same time. The two particularly chosen tests would measure; speed, agility and endurance. These tests were low cost, effective and efficient to conduct and repeat. This is so that anyone who desires to evaluate their athletes or referees can do so with relative ease and minimal cost with quality results.

**Test variables**

In this descriptive cross sectional analysis, we had two major variables, performance results for T-test and Yo-yo IR II test. The times resulting from the T-test and the distances from the Yo-yo II test were used as variables. We evaluated four sets of groups; KSÍ, HSÍ, and two semi pro teams, one handball team and one football team. The emphasis was on the Football referees and their performance.
The T-Test

The T-test is a simple and effective agility test designed to test athletes in their agility in forward, lateral and backward running.

![Diagram of T-test setup]

Figure 1. T-test. Derived from (Wood, 2008).

The procedure: The T-test was set up on a flat dry plane. It was conducted indoors on a court so the conditions would always be the same and the rate of injury risk is much lower due to dry conditions (Bangsbo, et al., 2006). The equipment necessary to perform the test is; 20 meter measuring tape, at least 4 cones, a set of timing gates, and an athlete to participate in the test. The cones were arranged in the shape of the letter T (see figure 1). This set-up, if viewed from above, looks like the letter T with gates at the bottom and the three objective cones at the top of the T.

The first part of the procedure is cone A, where the timing starts. The first cone of three that the athlete must touch is cone B which is placed 9.14 meters in front of cone A. The next two cones, cones C and D, are to be placed 4.57 meters on both adjacent sides of cone B. The reasoning for the strange distance measurements between cones is due to the conversion of a standard measurement to metric. The test was originally developed in America on the distances of 10 yards (9.144 meters) between cones A and B and 5 yards (4.572 meters) from cone B to either of the cones on the sides. The objective is for the athlete to sprint from the starting point to cone B. The athlete must touch each cone with the same hand. After touching cone B the athlete must side shuffle to the next cone, touch the cone with the same hand as before then
side shuffle over to the other outside cone and touch it again with the same hand. Then finally side shuffle back to cone B and touch it and turn towards the start line and sprint to the finish. The conductor will take time it takes the athlete to run to all the cones and return through the start/finish line.

**Rules:** It was ensured that the subjects face forward when shuffling side to side. The athletes were told not cross their feet when doing the side shuffle. This was to ensure measurement of agility speed and not just speed. The reason for this is that during a match the assistant referee more so then the main referee will find himself side shuffling. If he is prone to side shuffling and crossing his feet he increases the risk of entangling his feet and falling over, something no one wants to experience with live audience or during a broadcasted game (Salvo, et al., 2011).

**The Yo-yo intermittent recovery test level II**

The Yo-yo intermittent recovery test level II is a repeated sprint ability test that evaluates and determines an athlete's endurance as well as providing a way to calculate VO2max. There is a level 1 test for those whom are not as advanced of athletes. The main difference is that the test starts at 10 km/h vs the level 2 test which was used in this study begins at 13 km/h (Krstrup et al., 2003). The level 2 test speeds up gradually between each lap. Each lap is twenty meters with a 5 meter stopping area. The athlete’s objective is to listen to an audio file that is played allowed that beeps for start/return and stop counts. Said athletes must make it to the 20 meter line and back within the allotted amount of time, which gets progressively shorter as the speed of the test increases. The times for the shuttle runs increase with each round. As seen in the table below you can see the speed increases along with the number of laps that must be made. This is what makes this test hard enough to physically see what the endurance capabilities of the athletes whom participate in it actually are.

The actual demands of a Yo-yo test are highly comparable to activity demands of most sports. Short steady sprints with a change of direction. This is particularly so with football and team handball. The beauty of the Yo-yo test is that it allows researchers to develop results for an athletes’ ability to recover from a repeated exercise with a high utilization of their anaerobic system (Bangsbo, et al., 2008).
### Figure 2. Speeds and difficulty levels of Yo-yo II test.

<table>
<thead>
<tr>
<th>Speed Level</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>13</th>
<th>13.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speed Level</th>
<th>13.5</th>
<th>14.5</th>
<th>15.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep</td>
<td>5</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

### Figure 3. Yo-yo II Field Diagram.  
*Derived from (Lockie o.fl., 2017).*

**Necessary Equipment:** An indoor location with a track was chosen. Other equipment needed were plenty of cones for delineating distances, measuring tape to measure the appropriate distances for the test, the peep test recording to play for the athletes so they know the cadence to run on, a device to play the peep test on, and some form of ledger to record athlete’s times.

**The Procedure:** Two lines of cones were set to mark 20 meters of distance. The athletes started behind the front line of cones, they proceed to run crossed to the other set of cones/line when the audio recording instructs and return when the peep calls out to return. The speed of the peeps increases with every passing minute. The athlete must reach the line before the peep sounds. If they fail to reach the line before the peep then the athlete has two more peeps to catch up to the pace. If the athlete fails to do so then they are done with the test and their results for completed laps will determine their score.
The test was conducted on an indoor court so the conditions were always reparable. This is also done to minimize any risk of injury as the athletes were expected to be running at their full capacity un-interrupted from conditions that could present themselves when testing is administered outside, such as wet surfaces and cold weather, both of which are not advantageous for any athletes to perform if you want a relative measurement of their full capability due to heightened injury risk (Orchard and Powell, 2003).

**Scoring:** The score is the total distance covered by the athlete before he/she misses the peeps and cannot complete their objective in time or if they drop out of the test due to pure exhaustion. Testing times could last between 2-10 minutes.

**Statistical analysis**

The SPSS statistical program (version 24) was used to calculate mean, standard deviation and standard error of the mean (SEM). A t-test was used to calculate statistical difference between two groups with 95% confidence interval given. A p < 0.05 was considered statistically significant.
RESULTS

Comparison between KSÍ and HSÍ referees

This section describes the results obtained when 18 KSÍ referees and, for comparison, 19 HSÍ referees, when they were measured for physical endurance using the tests T-test and Yo-yo II test.

T-test results

The physical T-test was used in order to determine whether a difference was observed between the referee leagues from KSÍ and HSÍ. The mean time of 10.07 ± 0.19 seconds was observed for the 10 participants in the KSÍ referee group compared with 10.16 ± 0.15 seconds for the 19 participants in the HSÍ referee group (figures 4). When using an independent sample test the difference was not statistically significant ($t(27), -0.356, P>0.05$). This shows that the average KSÍ referee is 0.1 second faster on average than the average HSÍ referee.

Figure 4. T-test results for KSÍ and HSÍ referees.
Yo-yo II test results

In addition to the T-test, the two groups of referees also completed the Yo-yo intermittent recovery test level two, where the higher the score, the better the performance. When using an independent sample test there was not a significant difference between the two groups (t(26) = -0.936, p > 0.05). Both sets of referees had individual referees that ran much further than the others, but on average the handball referees showed more capability in the yo-yo II test in comparison to the football referees (figures 6). Eighteen KSÍ referees and 10 HSÍ referees participated in Yo-yo II endurance test. The mean distance for the football referees was 364 ± 34 meters, whereas the handball referees had a mean distance of 420 ± 52 meters (figures 5).

![Yo-yo II test](image)

*Figure 5. Yo-yo II test results for KSÍ and HSÍ referees.*
Comparison between HSÍ referees and semi-professional handball players

This section describes the results obtained when 19 HSÍ referees and, for comparison, 13 semi-professional handball players, were measured for physical endurance using the tests T-test and Yo-yo II test.

T-test results

When the two groups serving as comparison groups for the KSÍ referees, the HSÍ handball referees and the semi-professional handball players, were analyzed using T-test, the 19 HSÍ referees had a mean time of 10.15 ± 0.15 seconds and the 12 semi-professional handball players had a mean time of 9.96 ± 0.15 seconds (figures 6).

This is not a huge difference but nonetheless the semi-professional handball players performed significantly better than the HSÍ referees. We calculated the data with an independent sample test and it did not show a significant difference (t(29), -0.877, p>0.05). The interval differences were calculated with 95% confidence and it showed unremarkable results. Handball referees and handball players are very similar in speeds as far as the physical T-test is concerned.

Figure 6. T-test results for HSÍ referees and semi-professional handball players.
Yo-yo II test results

The two handball comparison groups were subjected to the level two Yo-yo intermittent recovery test where longer the distance covered the better the score. The independent samples showed that there was no significant difference between the two groups (t(21), -1.976, p>0.05), where the mean distance for the 10 handball referees was 420 ± 52 meters but 550 ± 42 meters for the 13 semi-professional handball players (figures 7). The interval difference for the handball referees was 234 meters, whereas for the semi-professional handball players it was 184 meters. Within the HSÍ referee group, their most endured referee ran 537 meters but their lowest conditioned referee ran 303 meters, whereas for the semi-professional handball players the most conditioned athlete ran 643 meters but their least conditioned (but still well in shape) ran 459 meters.

Figure 7. Yo-yo II test results for HSÍ referees and semi-professional handball players.
Comparison between KSÍ referees and semi-professional football players

This section describes the results obtained when 10 KSÍ referees and, for comparison, 18 semi-professional football players, were measured for physical endurance using the T-test.

T-test results

The secondary analysis with the KSÍ referees was against the semi-professional football players. There was a significant difference between the two groups. The mean score for the KSÍ referee group was 10.07 ± 0.19 seconds compared with 8.76 ± 0.08 seconds for the semi-pro football player group (figures 8). There was a significant difference between the two groups when we calculated the results with an independent sample test (t(26), 7.374, p<0.05). The results show with 95% measurable confidence that the football players showed a two full seconds faster completion of the T-test than the KSÍ referees. The KSÍ referees had an interval of 0.85 seconds with a 95% confidence in interval difference, whereas the football players had an interval of 0.34 seconds with 95% confidence in interval difference.

![T-test graph](image.png)

Figure 8. T-test results for KSÍ referees and semi-professional football players.
DISCUSSION

This study was conducted to evaluate the physical endurance of football referees belonging to KSÍ, and also to determine whether the tests used in the study were appropriate for measuring not only endurance, but also agility and anaerobics. The tests chosen were the T-test and Yo-yo II test.

The first test we conducted was the T-test. We administered this upon both groups of referees on different dates. KSÍ performed a little better than the HSÍ referees by a tenth of a second. This was not what we had expected since the HSÍ referees were in the middle of their regular season but the KSÍ referees coming off a winter break and starting their off-seasons training. This could be due to some of the differences in the sports they officiate in. There is vast difference between the required movements during game time. Handball referees cover a much smaller court than the football referees. The pitch size (diagonal distance crossed the main area of the court) and the speed of the sport in action contributes to this fact. A handball pitch is exactly 40 meters in length and 20 meters in width. A football pitch size can be a maximum of 120 meters in length and 90 meters in width. When compared, the recommended movement of a football referee is between the two penalty areas since he has help from his assistant referees. That length is roughly 80 meters. Football referees must always be moving to keep up with the ball whilst handball referees are more prone to position themselves, one behind the offensive line and one close to the end-line of the field. The handball referees work in couples allowing them to divide the court meaning they only need to move around 25 meters at most while the main ref of football is alone on the pitch with two assistants on either side of the field. The size of the area covered on average by the two types of referees is, therefore, notably different. This could have been a variable that played into the results of the T-tests. It may be concluded that the overall distance covered by KSÍ referees during games render them with more endurance as measured by the T-test, even though they are in off-seasons training.
The two referee groups were compared using Yo-yo II test. In the T-test there were 10 KSÍ referees and 19 HSÍ referees that participated, whereas in the Yo-yo II test the numbers were reversed with 18 KSÍ referees and 10 HSÍ referees participating. We found that the HSÍ referees covered a significantly longer distance than the KSÍ referees. This was not in agreement with the results of the T-test. The HSÍ referees were measured late in January when the regular season of the handball league is well underway. The KSÍ referees were measured in February, at a time when they had only recently started working out full time again with the pre-season competitions starting in early February. This indicates that the KSÍ referees are being measured while in their pre-season form while the HSÍ referees should be in top form. Knowing that the referees were measured at different times during the sport season, we had to be careful when drawing conclusions from the results. Because these two sports are not played during the same time of the year, it is impossible to measure both groups simultaneously. In order to obtain as reliable data as possible, both groups should be measured at the same times of their seasons, e.g. at the beginning, middle and the end. This would also add information about how they are progressing during the match seasons. In addition, it could provide data about the level of fitness of Icelandic referees in comparison to the European standards. We could use the results to further our development of what would be beneficial to the training of the referees to be at the top of their game. It would be astounding to see a league of Icelandic referees that are in the best shapes of their lives vs just being at an adequate/ acceptable standard of fitness.

When we calculated the 95% confidence interval of the Yo-yo II test it was interesting to see that the lower bracket was quite close to each other for the referee groups, i.e. HSÍ with 303 meters and KSÍ with 293 meters. That was not the case with the upper bracket where the HSÍ referees were significantly higher than the KSÍ referees, which suggests some of their participants scored very well on the test. This shows us how vastly different the individuals within the HSÍ group were as the referees ranged from poorly fit to very fit. The KSÍ referees, coming off a break, were still relatively similar, some being in poor condition as seen for some the HSÍ referees, but none in excellent physical form.
With this knowledge we conclude that the handball association must enforce stricter guidelines for their referees, it was visible how out of shape some of them were and the worst performance of all participants in the Yo-yo II test was from one of the handball referees in the middle of his season. It may be worrying for the HSÍ how bad of shape some of their referees were in. HSÍ should maintain a better fitness requirements, as a referee who is out of shape, although qualified to officiate a match, could be subject to making poorer decisions during tough situations. In addition, a referee who is visibly out of shape may present a negative image and be subjected to negative scrutiny.

These two tests used in this study were chosen because they are relatively simple to conduct, administer and evaluate when measuring a multitude of athletes accurately at the same time. The two particularly chosen tests would measure speed, agility and endurance. These tests were low cost, effective and efficient to conduct and repeat. This is also beneficial as anyone who desires to evaluate their athletes or referees with these tests can do so with relative ease and minimal cost with quality results.

We also compared the two referee groups with the players they officiate. As was to be expected, the athletes performed better than the referees on all accords of testing that were conducted. It may be concluded that this could be due to the fact that the athletes play the entire field with much more intensity than a referee. In addition, they are on average much younger than the referees, as it has been shown that on average the players are about 15 years younger than the referees. A study by Castagna et al. showed that age negatively affected the physical performance of the referee. The test conducted in that study did not include the testing we performed, i.e. T-test and Yo-yo II test. However, it would be interesting to further our research and see if there is considerable correlation between results (Castagna, et al., 2005).

We soon discovered the difficulties induced by the different season of each sport. Handball in Iceland is played from September until Christmas when they take a four to six week break and they start again in January to complete the regular season in March. The playoffs are then held between March and late May. Therefore, most of the referees need to be in shape from September and at least until March. Then the
playoff referees continue their program while others may get an early summer vacation from officiating. Football in Iceland is almost completely opposite to this, the season starts in late April or early May and is played throughout the Summer and ends with the final round being played in late September or early October. The reason for this is the fact that football is played outdoors and the majority of teams still play on old fashioned grass turf. There is change coming to get more teams onto artificial turf but some clubs are more rigid then others. The weather in Iceland does not permit grass fields to grow until around April and even then, the temperature can go below freezing which is detrimental to the grass. This has been seen with the 2018 season, which just started as this is being written and already in the first round two teams switched home-games so that the match would be played on artificial turf and not on a poor natural grass pitch.

We did meet the football referees more than once and their attitude and spirit seemed to be more positive at later appointments compared with the first appointment (personal evaluation). In addition, we felt that some of the referees showed improvements as time passed and their training became more difficult; however, this was not substantiated by data. Also, we felt that the football referees overall showed more athleticism then the handball referees, e.g. in physical appearance. Whilst the handball referees did in fact do better in the Yo-yo II test we feel as if the difference should have been greater as they were in their mid-season.

We would have liked to further the research and develop our understanding of all that encompasses to be a referee. For instance, we would have liked to measure for VO2 max. We had not organized the equipment needed to do laboratory tests, which could have shown the V02 max values of each referee. This would have been very useful to determine each individual referee’s level of fitness, but we did not have the necessary equipment prepared.

The analysis and comparison of the two referee groups gave an understanding of where the capabilities of the referees should be to officiate the events with utmost capability. The analysis and testing of these groups presented us with the results needed to show where a referee’s abilities should be and hopefully set a standardized level of physical requirement. This standardization will allow the administering refereeing
agency (KSÍ) to provide top quality referees to officiate the matches, both domestically and foreign.

A few things could have gone better. To begin with, the individuals of the referee group did not seem very interested to be tested when we first measured the KSÍ referees in December 2017. They got a few days notice of our arrival, at that time they were meeting up once a week for strength and endurance exercises, which were overseen by our instructor in this thesis. The day we came to administer our test few showed up, suggesting that knowing they had endurance tests on their schedule, many opted out of practice. We conducted the tests that day and informed the ones that did show up to not be afraid of participating even though they were in their off-season. For many it can be difficult to be tested when not well prepared, this and the fact that physical tests can be very difficult, is our reasoning behind the low number of participants that first day. Overall, we would have liked to have had more participants in each group. It became clear that it was hard to determine how much the lack of numbers was influencing the outcome and other aspects. And the last thing was something we have mentioned already, the two main groups of this study, the football referees and the handball referees, have two very different seasons. It became evident that the results would not be as optimal as possible. However, we believe this study can be a pioneering project which can be expanded with additional testing done in the future.
CONCLUSION

In this study, we used two tests to evaluate KSÍ referees in their skills. The main findings show that even though the KSÍ referees are out of season in their training schedule, they are in better shape when using the T-test. However, the HSÍ referees were able to run longer distances as measured by the Yo-yo II test, which coincides with their testing being performed in their mid-season training schedule. As the KSÍ referees performed better in one out-of two tests, outside their training season, it may be concluded that they are in better overall shape. This needs to be confirmed in a larger study performed during same points of time regarding seasonal training for both groups.

To further and enhance the referees’ capabilities as athletes in the areas of strength, agility, speed and endurance, we developed a series of instructional workout videos. See attachment below for the web addresses. The premise of the videos is to utilize these workouts as a standard base for the maintenance shape a referee should hold. Each video is short but precise. We took all things into consideration in the compilation of these videos. How the referees need to have a good amount of flexibility with explosive strength, both that can be utilized in the working environment on and off the field. We even went ahead and provided some instruction on how to reduce the impact of having to travel long distance before a match.
CITATIONS


APPENDIX

This appendix contains links to training videos for athletes with major emphasis on football referees and their physical mobility.

Lunges
https://youtu.be/FkEUFtgkgdU

Calf raises
https://youtu.be/BZGs-nbXeuQ

Hamstring stretches
https://youtu.be/80dFEEn1URQE

Bulgarian Split squat
https://youtu.be/B9HO3P7wous

Glute stretch
https://youtu.be/u8I4hPLirxs

Quad stretch
https://youtu.be/VX34pV5DA4s

Dynamic hip stretch
https://youtu.be/4qote3EX-H4

Glute Bridge Exercise
https://youtu.be/ng281JeAO4o
Butterfly stretch/ groin stretch
https://youtu.be/JYLiUu6knUObk

Push up / plank
https://youtu.be/DEsIIv-YJp8