UNIVERSIDADE
BEIRA INTERIOR

Desempenho e avaliação dos árbitros e observadores dos árbitros de futebol

Vítor Manuel Pereira de Carvalho

Tese para obtenção do Grau de Doutor em
Ciências do Desporto
(3º ciclo de estudos)

Orientador: Bruno Filipe Rama Travassos
Co-orientador: Prof. Doutor Pedro Tiago Esteves

agosto de 2021

Nada é fácil, tudo é difícil.

Nada é difícil, tudo é fácil.

Enfrentar a Vida,

Dir-nos-á o fácil e o difícil.

O gostar de,

A motivação para,

O desespero,

A frustração,

A alegria,

A solidão,

O regozijo,

O desencanto,

A persistência,

O alcançar do objetivo.

Sucesso, é tentar...

Aos meus Avós

Agradecimentos

A conceção de uma Tese de Doutoramento é indiscutivelmente uma ação desenvolvida ao longo de um determinado período de tempo, gerando a envolvência e interação de um conjunto diverso de pessoas sem as quais as dificuldades teriam sido superiores no almejar, no alcançar dos objetivos propostos.

Manifesto o meu agradecimento a todos aqueles que direta e indiretamente estiveram conectados, contribuindo de forma altamente positiva e dinâmica para a boa persecução e apresentação do presente trabalho desenvolvido.

- Ao Professor Doutor Bruno Travassos, meu orientador, as palavras são inócuas para revelar o quanto importante foi em todo o caminho. Ressalvo a competência, a honestidade e a clarividência, colocados em prática, naqueles momentos em que a complexidade transcendia a realidade.
- À Professora Doutora Célia Nunes, a complexidão da estatística versus a sua persistência em busca da solução que desbloqueasse o problema. A competência, a colaboração, a disponibilidade no momento em que a necessidade imperava.
- Ao Professor Doutor Pedro Esteves, ressalva-se a competência o rigor nos momentos da definição do caminho a prosseguir demonstrando vasta disponibilidade e estímulo.
- Ao Conselho de Arbitragem da Federação Portuguesa de Futebol (2016-2020), pelo acesso à documentação disponibilizada para a boa prossecução deste trabalho.
- Aos senhores observadores, pertencentes ao quadro de observadores de elite na época desportiva de 2016-2017, pela enorme disponibilidade, atenção e dedicação a este trabalho, com prejuízo pessoal, mas com um enorme sentido do dever e de colaboração.
- À minha filhota, sempre presente, *Carpe Diem … Dady …*

Resumo

A presente tese teve como objetivos: (i) compreender o nível de concordância entre diferentes métodos de observação e avaliação do desempenho dos árbitros de futebol realizada pelos Observadores dos Árbitros; (ii) avaliar a relação entre os indicadores de desempenho dos árbitros de futebol e a sua avaliação num jogo de futebol; (iii) avaliar a influência de variáveis contextuais e situacionais na classificação dos árbitros de futebol de elite no decorrer de uma época desportiva em Portugal. Os resultados revelam relativamente ao primeiro objetivo: a) a existência de diferentes níveis de concordância no julgamento e avaliação do desempenho do árbitro face aos diferentes métodos observados; b) evidências que o método de avaliação ao vivo seguido de observação em vídeo parece ser o método mais ajustado para a melhoria da fiabilidade da observação realizada. Relativamente ao segundo objetivo: c) as variáveis, aplicação negativa das leis do jogo e trabalho em equipa, explicam cerca de 60% da avaliação do desempenho do árbitro de futebol. Na concretização do terceiro objetivo os resultados demonstram que: d) as variáveis contextuais revelam um efeito significativo sobre a classificação final, não se verificando efeitos significativos das variáveis situacionais, sobre a classificação dos árbitros, no final da época desportiva. Os resultados obtidos permitem uma melhor compreensão sobre o processo de avaliação do árbitro de futebol e das implicações das diferentes metodologias adotadas para a sua classificação. Mais pesquisas devem ser desenvolvidas, para compreender o impacto dos métodos aplicados, assim como identificar com maior precisão qual a influência das variáveis contextuais na avaliação/classificação no final do jogo, bem como no final de uma época desportiva, de um árbitro de futebol, com a finalidade de promover uma melhoria de desempenho e desenvolvimento da carreira do árbitro.

Palavras-chave

arbitragem; avaliação; observação jogo; tomada de decisão; indicadores de performance

Abstract

This study aimed: (i) to understand the level of agreement between different methods of observation and assessment of the performance of football referees carried out by the Referee Match Observers; (ii) to evaluate the relationship between the football referees' key performance indicators and their assessment at a football match; (iii) to evaluate the influence of contextual and situational variables on the classification of elite soccer referees during the sports season in Portugal. The results revealed in relation to the first objective: a) different levels of agreement between methods used in the judgment and assessment of the referee's performance; b) live assessment method followed by video observation seems to be the most adjusted method to improve the accuracy of the observation performed. Regarding the second objective: c) the variables negative application of the laws of the game and teamwork, explain about 60% of the game referee's assessment score at the end of each match. In the achievement of the third objective, the results revealed that: d) the contextual variables, revealed a significant effect on the final classification at the end of the sports season. In opposition, the situational variables didn't reveal any effect on the referee's classification at the end of the sports season. The results obtained allow a better understanding of the soccer referee's performance and assessment with implications on the methodologies adopted for referee's classification. Further research must be developed, to understand the impact of the methods of analysis used, as well as, to identify with greater precision the influence of contextual variables in the evaluation/classification of referees at the end of the game and the sports season, with the purpose of promoting an improving the performance and development of the referee's career.

Keywords

referee; evaluation; match observation; decision-making; key performance indicators

Publicações

Os seguintes artigos foram realizados como parte dos trabalhos de doutoramento:

- Carvalho, V., Esteves, P. T., Nunes, C., Araújo, D., Helsen, W. F., & Travassos, B. (2020). Observe and make a call: Football referee's assessment is context sensitive. *International Journal of Performance Analysis in Sport, 20(6), 982-993*. https://doi.org/10.1080/24748668.2020.1820194

- Carvalho, V., Esteves, P.T., Nunes, C., Costa, P., Helsen, W.F., Travassos, B. (Under revision). Identification of key predictors of Football referee's assessment: A case study in Portuguese league.

- Carvalho, V., Esteves, P. T., Nunes, C., Mendez, C., & Travassos, B. (2021). A influência de variáveis contextuais e situacionais na classificação de árbitros de futebol de elite. Cuadernos de Psicología del Deporte, *21*(1), 216–224. https://doi.org/10.6018/cpd.435081.

Foram ainda realizadas as seguintes apresentações em congressos técnico científicos:

- **Carvalho, V.**, Nunes, C., & Travassos, B. (2018) Fiabilidade na avaliação de árbitros de futebol in-loco e em vídeo. XIX Jornadas da Sociedade Portuguesa de Psicologia do Desporto. Edições ISMAI, 39.
- **Carvalho, V.**, Nunes, C., Esteves, P., Araújo, D., & Travassos, B. (2019) Evaluation of referees' performance in-loco and in video. Motricidade, 15: Sup1, 159. https://doi/0.6063/motricidade.16967.
- Travassos, B., **Carvalho, V.**, Nunes, C., Esteves, P. (2019) How in-loco and video analysis influences referees's evaluation. Internacional conference on football refereeing. Federação Portuguesa de Futebol.
- **Carvalho, V.**, Esteves, P.T., Nunes, C., Araújo, D., Helsen W.F., & Travassos, B. (2020): Observe and make a call: football referee´s assessment is context sensitive, Formação de Observadores de Futebol de 11, Federação Portuguesa de Futebol

- **Carvalho, V.**, Esteves, P.T., Nunes, C., Araújo, D., Helsen W.F., & Travassos, B. (2020): Observe and make a call: football referee´s assessment is context sensitive, Formação de Observadores de Futsal, Federação Portuguesa de Futebol

Índice

Agradecimentos v

Resumo vii

Abstract ix

Publicações xi

Lista de Tabelas xv

Lista de Acrónimos xvii

Capítulo 1 – Introdução Geral 1

Capítulo 2 – Objetivos 7

Capítulo 3 – Observe and make a call: Football referee's assessment is context sensitive 9

Capítulo 4 – Identification of key predictors of Football referee's assessment: A case study in Portuguese league 21

Capítulo 5 – A influência de variáveis contextuais e situacionais na classificação de árbitros de futebol de elite 31

Capítulo 6 – Discussão Geral 42

Capítulo 7 – Estudos futuros 49

Capítulo 8 – Referências bibliográficas 50

Anexos 56

Lista de Tabelas

Table 1 - Description of the analyzed KPI for each main criterion considered 13

Table 2 - Mean values of the criteria general evaluation for L, LV and V and Intra and Inter-observer analysis 15

Table 3 - Results of the KPIs for difficulty and intervention for L, LV and V and correspondent comparisons 16

Table 4 - Mean values of the criteria disciplinary actions and game management for L, LV and V and correspondent comparisons. 17

Table 5 - Description of the analyzed KPI for each main criterion considered. 24

Table 6 - Results of the KPI for difficulty and intervention for LFS and HFS. 26

Table 7 - Results of the KPI for criteria disciplinary actions and game management for LFS and HFS. 26

Table 8 - Estimated regression coefficients from binary logistic regression analysis, only for significant variables (final score as dependent variable). 27

Tabela 9 - Variáveis contextuais e situacionais que influenciam a classificação final dos árbitros de futebol. 37

Lista de Acrónimos

AUC	Area Under the Curve
CI	Confidence Intervals
CL	Confidence Limits
FA	Football Association (Federação Portuguesa de Futebol)
FIFA	Federação Internacional de Futebol
G1	1º tercil – Classificação superior (do 1º ao 8º, inclusive)
G2	2º tercil – Classificação intermédia (do 9º ao 16º, ambos inclusive)
G3	3º tercil – Classificação inferior (do 17º ao 23º, ambos inclusive)
HFS	Higher Final Score
IBM SPSS USA	Software para análise de dados
ICC	Intraclass Correlation Coefficient
K	Kappa de Fleiss'
KPI	Key Performance Indicators
L	Live (ao vivo)
LFS	Lower Final Score
LIGA NOS	Campeonato de futebol profissional 1ª categoria: Portugal
LV	Live + Vídeo (ao vivo + vídeo)
OA	Observadores de árbitros de futebol
OR	Odds Ratio
ρc	Lin's Correlation Coefficient of Agreement'
RMO	Referee Match Observers
ROC	Receiver Operating Characteristic
SD	Standard Deviation
UEFA	União das Associações Europeias de Futebol (Europe's Football Governing Body)
V	Vídeo (Apenas em vídeo)

Capítulo 1 – Introdução Geral

No futebol, o árbitro e os árbitros assistentes têm por objetivo a aplicação das Leis do Jogo (Catteeuw, Helsen, Gills & Wagemans, 2009). No entanto, aos árbitros de futebol não é apenas exigido que tomem decisões e apliquem a punição adequada (Elsworthy, Burke, Scott, Stevens & Dascombe, 2014), de forma fundamentada nas Leis do Jogo (Catteeuw et al., 2009), mas que o façam também de modo a garantir uma adequada gestão do jogo (Mascarenhas, Collins & Mortimer, 2002). Ao longo dos últimos anos, vários estudos avaliaram a prestação do árbitro, tendo por base a avaliação da sua tomada de decisão (Elsworthy et al., 2014), sendo, durante muitos anos, considerada a competência mais importante de um árbitro de futebol (Helsen & Bultynck, 2004; Morris & O'Connor, 2016). De acordo com este pressuposto, Helsen e Bultynck (2004), concluíram que os árbitros tomam aproximadamente 200 a 250 decisões por jogo, sendo observáveis cerca de 137. Destas, cerca de 28%, estão relacionadas com a Lei 12 (Faltas e incorreções) e cerca de 31%, estão relacionados com a Lei 15 (O lançamento lateral). Em cada uma destas decisões, o árbitro tem de avaliar, de forma rápida e ajustada às exigências do contexto, se assinala "falta/ falta e exibe cartão/ não exibe cartão", existindo a incerteza de uma decisão, que é condicionada pelo tempo e por um insuficiente acesso à possível informação relevante (Lex, Pizzera, Kurtes & Schack, 2015). Portanto, apesar do pouco tempo disponível e muitas das vezes de limitada informação, cada decisão e intervenção é evidentemente crucial, e um único erro do árbitro (penalti, fora de jogo duvidoso) pode afetar decisivamente o resultado do jogo (Halsey, 2007). Assim, a capacidade de decisão e de intervenção no jogo foi avaliada como um meio eficaz para diferenciar árbitros peritos de árbitros não peritos (Berry, Abernethy & Côté, 2008; Williams & Ericsson, 2005), considerando-se que refletia o nível de adaptabilidade do árbitro à variabilidade do próprio jogo (Kittela, Larkina, Elsworthy & Spittle, 2019; Urra, Nuñez, Oses & Sarmiento, 2018).

Características de um árbitro de excelência

Apesar do foco na tomada de decisão, o conhecimento sobre as características que distinguem um árbitro de excelência de um árbitro regular deve ir mais além, no sentido de uma maior compreensão sobre os motivos que estão a condicionar a tomada de decisão e a sua intervenção no jogo. Numa tentativa de compreensão multidimensional, Mascarenhas, Collins e Mortimer (2005) identificam cinco caraterísticas: a) conhecimento e aplicação da lei; b) julgamento contextual; c)

competências ao nível da personalidade e da gestão do jogo; d) características psicológicas de excelência e finalmente e) aptidão física, posicionamento e qualidades mecânicas. Slack et al., (2013) apontam para uma multiplicidade de fatores físicos, ambientais e psicológicos que contribuem para a excelência de desempenho do árbitro, como a resistência mental, as redes de serviços de apoio e a qualidade de gestão do jogo ou um superior desempenho físico. Por sua vez, Schnyder e Hossner (2016) relatam, como principais características, a educabilidade, a inteligência futebolística e a experiência do árbitro. Posteriormente, Pina, Passos, Carvalho e Maynard (2019) procuram sintetizar a investigação anterior, reforçando que a excelência do árbitro de futebol é um processo que decorre do desenvolvimento de três dimensões distintas, mas que concorrem entre si para o desenvolvimento da excelência na arbitragem: preparação individual, preparação do jogo e gestão do jogo.

De acordo com o referido, estudos anteriores revelam que os árbitros de elite não apresentam apenas estratégias antecipatórias mais eficazes que sustentam a sua tomada de decisão (Mather, 2008), mas também patenteiam uma melhor condição física, um menor nível de stress e tendem a assinalar menos infrações do que árbitros de categoria inferior (Guillén & Feltz, 2011), demonstrando a sua personalidade e capacidade de gestão do jogo, com consequentes melhorias na sua prestação final (Bar-Eli, Plessner & Raab, 2011).

Num estudo realizado por Helsen, Starkes e Hodges (1998) verificam que os melhores árbitros são os que revelam uma maior capacidade para aplicar corretamente os critérios disciplinares descritos nas leis do jogo, mas também para os aplicar de acordo com os requisitos do jogo. Fruchart e Carton (2012) sustentam esta ideia, reforçando que a gestão dos aspetos disciplinares por árbitros de elite, em comparação com árbitros de menor experiência, nomeadamente na exibição de cartões amarelos ou vermelhos, são influenciadas por "regras não escritas" (Plessner & Raab, 1999) permitindo uma gestão do jogo mais eficaz e de acordo com o contexto em que este ocorre (Askins, 2001; Grunska, 2001), em detrimento do estrito cumprimento das leis de jogo (Weaver, 2001). Nesta perspetiva, os árbitros de elite não são os que dominam apenas as leis de jogo, mas os que as aplicam de forma coerente ao contexto e dinâmica especifica, ajustando o seu julgamento e a sua intervenção à dinâmica do próprio jogo (Collina, 2003).

Para potenciar a sua decisão, é fundamental que os árbitros sejam dotados de grande capacidade percetiva de modo a que sustentem a sua tomada de decisão em informações relevantes sobre os jogadores, a bola e os árbitros assistentes de forma conjunta e antecipativa (Ghasemi, Momeni, Rezaee & Gholami, 2009). No entanto,

fruto da dinâmica do próprio jogo, os árbitros muitas vezes tomam decisões sem a certeza de um comportamento observável (toma a decisão de não intervir no jogo), evidenciando-se assim a inexistência de critérios objetivos na tomada de decisão, em consequência de uma limitada informação face à dinâmica e contexto do jogo (Weston et al., 2012).

Apesar do limitado conhecimento sobre os mecanismos percetivos (Weston et al., 2012) que sustentam as decisões tomadas pelos árbitros, Pietraszewskia, Maszczyka, Rocznioka, Gołaśb e Stanula (2014) revelam que as habilidades percetivas dos árbitros variam dependendo da função e do nível de realização profissional, sendo que a qualidade da sua perceção pode influenciar o número e a precisão das decisões. Do mesmo modo, a experiência e a idade do árbitro podem também determinar fortemente as suas habilidades percetivas. Num estudo envolvendo a comparação entre árbitros e árbitros assistentes, Catteeuw et al., (2009), relatam que os anos de arbitragem, as horas de prática por semana e o número de jogos arbitrados foram positivamente relacionados com o desempenho dos árbitros na tomada de decisão.
Aoba, Yoshimura, Miyamori e Suzuki (2011) identificam ainda que os árbitros mais experientes não utilizam apenas a melhor informação sobre o jogo, mas também procuram utilizar e compartilhar mais informações com os árbitros assistentes em comparação com os menos experientes. Para isso, eles reduzem a distância de movimento entre si, e melhoram o posicionamento de modo a manter um campo visual que permita visualizar todo o campo de jogo, na qual ocorre o jogo de forma conjunta. Helsen e Bultynck (2004) indicam a este propósito que 64% de todas as decisões dos árbitros de elite são baseadas no trabalho em equipa de arbitragem.

Tendo por base as exigências da atividade de arbitragem, bem como o contexto em que estas se desenvolvem, Mascarenhas et al., (2005) verificam que a performance do árbitro resulta da sua capacidade individual para intervir e gerir não só os fatores contextuais que marcam o contexto de jogo e o ambiente que o rodeia, mas também os fatores situacionais determinados pela dinâmica do jogo, assim como os aspetos disciplinares que o constituem. Curiosamente, os melhores árbitros não sofrem de forma significativa a influência de fatores contextuais e situacionais na sua perceção e tomada de decisão no jogo de futebol. A verdade é que para melhorarmos a perceção sobre o desempenho de um árbitro necessitamos de uma melhor compreensão sobre quais são as variáveis que mais influenciam a prestação do árbitro de forma genérica (Buraimo, Simmons & Maciaszczyk, 2012; Catteeuw et al., 2009). A respeito da influência do contexto no julgamento e tomada de decisão do árbitro de futebol, pesquisas anteriores indicam, que o seu desempenho é condicionado pelo efeito do

ruído da multidão (Garicano, Palacios-Huerta & Prendergast, 2005; Nevill, Balmer & Williams, 2002), pela vantagem de jogar em casa (Sutter & Kocher, 2004), pela reputação da agressividade das equipas (Jones, Paull & Erskine, 2002) pelas vocalizações dos jogadores (Lex et al., 2015), pela dinâmica do jogo (Unkelbach & Memmert, 2008), pelo tempo do jogo (Helsen, Gilis & Weston, 2006), e ainda pela combinação de fatores contextuais (qualidade da equipa) e situacionais do jogo (diferença de pontos, número de faltas) (Lago-Peñas & Gómez-López, 2016).

Por exemplo, Nevill et al., (2002) revelam que, os árbitros, que ouviam o barulho do público num clip de vídeo, eram mais influenciados na decisão sobre uma possível situação de infração da equipa da casa, do que os árbitros que assistiram aos clipes sem som ambiente. No entanto, embora fatores contextuais (equipa da casa ou público) pareçam influenciar a tomada de decisão do árbitro em competições nacionais (Buraimo, Forrest & Simmons, 2010; Buraimo et al., 2012), o mesmo não se verifica em competições internacionais, com árbitros de elite. De modo semelhante, parece que o nível competitivo das equipas influencia a gestão e preparação do próprio jogo (Tenreiro-Gavela, Montero-Seone & Saavedra-García, 2016), bem como o modo como os árbitros decidem a atribuição de um penalti (Erikstad & Johansen, 2020). Por exemplo, Jones et al., (2002) verificam que o conhecimento prévio dos árbitros, acerca da agressividade das equipas, influencia a exibição de cartões amarelos.

Por outro lado, relativamente aos fatores situacionais, o modo como o árbitro realiza a gestão do jogo pode levar a que uma mesma tomada de decisão ou uma mesma ação disciplinar possam ser interpretadas como um erro de julgamento face às leis de jogo, ou pode ser vista como uma decisão correta, de modo a salvaguardar a gestão e fluidez do jogo (Askins, 2001; Grunska, 2001). Plessner e Betsch (2001) revelam que a decisão de assinalar um penalti poderá ser influenciada por decisões anteriores em situações similares, já tendo sido demonstrado em outros contextos desportivos que as “regras não escritas” podem influenciar as decisões dos árbitros (Plessner & Raab, 1999). De forma semelhante, Lex et al., (2015) revelam ainda que a ação disciplinar do árbitro através da exibição de cartão amarelo era mais vezes registada perante a vocalização do jogador e pela necessidade de gestão do jogo, do que propriamente fruto de uma infração às leis de jogo.

Avaliação do árbitro de futebol

Apesar do referido anteriormente, o desenvolvimento das competências de excelência do árbitro são altamente influenciadas pelo processo que as medeia, nomeadamente, as classificações, promoções e as despromoções de categoria (Fruchart & Carton, 2012), assim como o julgamento e a avaliação dos Observadores de Árbitros (OA), tendo em consideração cada contexto específico do jogo (Pina et al., 2019).

A respeito da avaliação da prestação do árbitro de futebol e consequentemente da classificação no final da época desportiva, Weaver (2001) considera que este processo influencia bastante a prestação do árbitro no decorrer do jogo, não permitindo em algumas circunstâncias adotar uma gestão do jogo mais eficaz e de acordo com o contexto em que este ocorre, em detrimento do estrito cumprimento das leis de jogo. As avaliações do desempenho do árbitro são normalmente efetuadas por ex-árbitros experientes, que de acordo com as diretrizes da Federação Internacional de Futebol (FIFA) e da União das Associações Europeias de Futebol (UEFA) e do próprio conselho de arbitragem da federação nacional, avaliam e classificam o desempenho do árbitro de futebol num determinado jogo. A avaliação do desempenho dos observadores de árbitros incorpora diferentes dimensões do desempenho de um árbitro (Mathers & Brodie, 2011; Pina et al., 2019; Russell, Renshaw & Davids, 2018). Esta avaliação é realizada tendo por base uma observação ao vivo ou em vídeo, baseada em grelhas de observação, integrando as seguintes categorias de observação, que têm por base os principais indicadores de desempenho do árbitro de futebol: i) dificuldade e intervenção no jogo, no qual se tem em consideração fatores contextuais, capacidade individual e capacidade de posicionamento bem como colaboração entre a equipa de arbitragem; ii) ações disciplinares e gestão do jogo, no qual se consideram aspetos relacionados com ação disciplinar, aplicação das leis de jogo e controlo disciplinar (ver Anexo 1).

Apesar deste processo de avaliação estar bem definido, e tendo por base regras muito claras e objetivas para os avaliadores, nada se sabe até ao momento sobre qual a relação entre cada indicador de desempenho avaliado nas grelhas de observação criadas para o efeito e a avaliação final da prestação do árbitro. Do mesmo modo, como o contexto influencia o julgamento do árbitro em campo bem como a tomada de decisão, também é possível que o ambiente do estádio, ou a observação dos árbitros através de vídeo restrinjam o julgamento e a avaliação dos árbitros pelo OA. Portanto, a concretização de processos de avaliação de árbitros de futebol em diferentes contextos de observação e com recurso a diferentes meios e perspetivas de avaliação pode causar divergência na

pontuação final do desempenho do árbitro no jogo (Halsey, 2007), fruto da restrição ou ampliação da informação tida em consideração para o processo de avaliação. Portanto, a possibilidade de os OA experienciarem diretamente ou não o desempenho do árbitro pode gerar implicações extraordinárias, relativamente aos procedimentos de observação e avaliação.

Considerando assim que a avaliação dos árbitros de futebol envolve múltiplas intervenções de diferentes agentes, é de extrema importância atingir elevados padrões de fiabilidade de análise. A este propósito, podemos referir que os OA assumem uma responsabilidade e uma exigência na sua atuação crucial para a classificação e melhoria do desempenho dos árbitros de futebol, consequentemente para o desenvolvimento da sua carreira (Fruchart & Carton, 2012; Mathers & Brodie, 2011). Portanto, torna-se indispensável desenvolver o conhecimento sobre os meios de observação, bem como sobre se as alterações aos métodos de observação ao vivo, em vídeo, ou a combinação de ambas, podem influenciar no julgamento dos OA.

Capítulo 2 – Objetivos

Face ao exposto anteriormente, o presente trabalho, tem como objetivo desenvolver uma melhor compreensão, sobre o processo de avaliação do árbitro de futebol e verificar a implicação ao recurso a diferentes métodos de avaliação, assim como das variáveis observadas e da influência de fatores contextuais e situacionais, na avaliação e classificação do árbitro de futebol, não apenas no jogo, mas também no final de uma época desportiva.

Deste modo, este trabalho tem como objetivos específicos:

- Compreender o nível de concordância entre diferentes métodos de observação e avaliação do desempenho dos árbitros de futebol realizada pelos Observadores de Árbitros de Futebol (OA).
- Avaliar a relação entre as variáveis de desempenho avaliadas pelos OA de futebol e a sua avaliação no final do jogo.
- Avaliar a relação entre a classificação dos árbitros de futebol de elite e as variáveis contextuais e situacionais dos jogos arbitrados.

O trabalho foi organizado por capítulos com o seguinte formato de apresentação:

No capítulo 1, realizamos um enquadramento do trabalho através de uma revisão da literatura sobre a caraterização das exigências do árbitro de futebol e os desafios à sua avaliação. Feita a revisão da literatura no capítulo 1, definimos os objetivos do trabalho no capítulo 2.
Tendo por base os capítulos anteriores, apresentámos nos capítulos 3, 4 e 5, os três estudos de investigação realizados. No referente ao capítulo 3, procura-se uma melhor compreensão sobre o nível de concordância, entre os métodos de observação do árbitro de futebol: i) ao vivo; ii) ao vivo + vídeo; e iii) apenas em vídeo, realizada por OA de elite. No capítulo 4, averigua-se por sua vez qual a relação, entre as variáveis de desempenho avaliadas pelos OA, nomeadamente, as respeitantes à dificuldade e intervenção no jogo e as ações disciplinares e gestão do jogo, e a sua avaliação no final do jogo. No capítulo 5, procura-se avaliar numa dimensão mais macro, qual a relação existente entre a classificação dos árbitros de futebol de elite no final da época desportiva e as variáveis contextuais e situacionais dos jogos arbitrados.

No capítulo 6, efetuamos uma discussão dos trabalhos realizados, bem como apresentamos algumas sugestões para a melhoria da fiabilidade do processo de avaliação do desempenho do árbitro de futebol e alvitrámos algumas aplicações práticas para o processo de avaliação. Por fim, no capítulo 7, são propostos os estudos futuros que daqui podem advir para uma melhor compreensão do processo de avaliação do árbitro de futebol.

Capítulo 3 – Observe and make a call: Football referee's assessment is context sensitive

Abstract

This study aimed to understand the level of agreement between different methods of observation and assessment of the performance of football referees carried out by Referee Match Observers (RMOs): i) live; ii) live + video and iii) single video. A total of 34 observers from the 1st category of the Portuguese Football Association (FA) participated in the study. The evaluation followed the official assessment guidelines according to three criteria: a) general evaluation; b) game difficulty and general performance; and c) disciplinary actions and game management. Results revealed different levels of agreement in the judgment and evaluation of the referee's performance between live, live + video and video only. Results suggest that the combination of live and video analysis allow RMOs to be more perceptually attuned to the events that convey relevant information to better support the assessment of referee's performance. This work offers evidence for the redesign of official observation procedures and policies of football governing bodies that are in charge in charge of referee's match evaluations.

Keywords: matches observation reliability; contexts of observation; match observers; judgement; decision-making

---

Carvalho, V., Esteves, P. T., Nunes, C., Araújo, D., Helsen, W. F., & Travassos, B. (2020). Observe and make a call: Football referee's assessment is context sensitive. *International Journal of Performance Analysis in Sport, 20*(6), 982-993. https://doi.org/10.1080/24748668.2020.1820194

Introduction

In football, the referee and the assistant referees aim to apply the laws of the game coherently in straight connection with the game dynamics (Catteeuw et al., 2009). In fact, the study of self-perception of football referees on their own decisions has disclosed the intention to maintain control and preserve the integrity of the game by assembling a competitive and cooperative relationship with the players under interacting constraints (Russell et al., 2018).

It is well documented that the context influences on-field referee judgement and decision making in association football. For instance, previous research has alluded to the effect of crowd noise (Garicano et al., 2005; Nevill et al., 2002), the advantage of playing at home (Sutter & Kocher, 2004), the reputation of the teams for aggressiveness (jones et al., 2002), or the player's vocalizations (Lex et al., 2015), along with the game dynamics (Unkelbach & Memmert, 2008), the game period (Helsen et al., 2006), or even the combination of contextual (team level) and situational game variables (score difference, number of fouls) (Lago-Peñas & Gómez-López, 2016). Recently, it has been advocated that the investigation of referee expertise should go beyond the single contextual effects by considering decisions as interdependent over time and nested within the context of the global event (Russell et al., 2018). In fact, one of the most challenging issues in football refereeing implies managing interpersonal relations with the players over the course of the game to guarantee coherence between actual and previous decisions (Mallo, Frutos, Juárez & Navarro, 2012). Also, given that the game changes on a moment-to-moment basis, the referee must permanently adjust their position in the field to ensure the best perspective to perceive the game incident and to make accurate decisions (Al Hazmi, 2016; Catteeuw et al., 2009). In addition, the referee should critically decide within a time window that allows to pick up relevant information about the players, the ball and the fellow assistant referees (Ghasemi et al., 2009; Helsen & Bultynck, 2004). The selection of such informational cues support proper decision-making and have been underlined as key performance indicators to achieve a high level of refereeing performance (Helsen & Bultynck, 2004). For example, it was reported that the distance to the ball, the angle of viewing and the velocity of the referee at the moment right before the decision are crucial to result in accurate decisions (Al Hazmi, 2016; Elsworthy, Burke & Dascombe, 2014).

The evaluation of football referees is operationalized by external Referee Match Observers (RMO) who are responsible for giving a mark after the game. The role of the RMO is itself very complex yet of crucial relevance not only for the classification of football referees but also for performance enhancement and career development

(Fruchart & Carton, 2012; Mathers & Brodie, 2011). Indeed, the final grade atributed to the referee at the end of the sports season directly influence their upgrade, downgrade or maintenance in a certain level (Fruchart & Carton, 2012).
Usually, the RMO go to the stadium in order to evaluate live the performance of the referees. Their analysis comprises the evaluation of the technical and disciplinary decisions, but also the difficulty of the game, the positioning, as well as the capacity of game management and collaboration with assistant referees (Mathers & Brodie, 2011). However, as the context influences the on-field referee judgement and decision making, also it is possible that the stadium environment or the analysis of referees through video also constraints RMO judgement and evaluation of referees.

Technological innovation and development in recent years, promoted the use of new technological devices in soccer, opening up the possibility to reduce errors of judgment and decision-making from the referee (Dohmen & Sauermann, 2016; Schweizer, Plessner, Kahlert & Brand, 2011), as well as the evaluation of their performance by RMO. Accordingly, it was argued that to improve the reliability of football referee's assessment, precise and reliable digital video technology should be used by RMO (Halsey, 2007). In fact, in the Referee Convention guidelines of Europe's Football Governing Body UEFA, it is mentioned that match observations have to be organized live in the stadium. As from the last season on, they introduced a new evaluation system that combines two different observations live and video by 2 independent RMOs. However, for instance, in the English Premier League, match observations are implemented just by the use of video. By keeping in line with the complexity of factors underpinning football refereeing (Dohmen & Sauermann, 2016; Pina et al., 2019), the possibility for the RMOs to directly experience or not the referee's performance could generate notable implications on their observation and evaluation procedures. Considering the fact that the evaluation of football referees involves multiple intervention of different agents, it is of foremost importance to achieve high standards of reliability. However, empirical evidence of this kind is still missing.

Previous research Put et al., (2016) involving football assistant referees in football compared the use of a three-dimensional (3D) and a two-dimensional (2D) simulations in an off-side decision-making task. Results showed a higher response accuracy under 3D compared to 2D task that indicate that the visual system can benefit from the availability of 3D information, especially in the case of complex and dynamic tasks such as football game. However, more knowledge is needed on whether changes in the methods of observation from *live*, to *video*, or the combination of both may influence on the judgement of RMOs. Crucially, this still open question can generate a high level

of impact on football referee's assessment (Gilis, Helsen, Catteeuw & Wagemans, 2008). To our best knowledge no investigations were made about the potential differences between football referee's assessment when using single *live* or *video* analysis or using the combination of live + video.

Accordingly, this study aimed to understand the level of agreement between the most usual method, live, and the other possible methods of observation live + video and using only video. Specifically, it was tested the level of agreement between RMOs that observed: i) live (L) with live + video (LV) (L vs LV); ii) live (L) with video (V) (L vs V) and between live + video (LV) with video (V) (LV vs V). It is our expectation to observe a lower level of agreement between L and V due to changes on context and on the perspective of RMO analysis. However, it is also expected that different variables revealed different levels of agreement between L, LV and V due to the different perspectives required to analyse different dimensions of performance.

Methods

Sample

Thirty-four elite Referee Match Observers (RMO) from the Portuguese (FA) voluntarily participated in the study (Age – 52.62 ± 8.95; Number of years as observer – 12.11 ± 6.81; Number of years as observer in 1st league – 7.18 ±5.69). An informed and written consent was provided to the Portuguese (FA) referee Council, as well as to all the participants, before the beginning of the study. All participants were informed about the goal of the study and notified that they could withdraw from the study at any time. The study protocol followed the guidelines stated in the Declaration of Helsinki and was approved by the Local Ethics Committee.

Data collection

Thirty-four games were observed by 34 elite RMOs using three contexts of observation: (a) live (L). L context implied that the observers perceived and analyzed the game in the stadium and submitted the evaluation within 24 hours to the Portuguese (FA), according to the protocol previously defined; (b) live + video (LV). LV context meant that the observers observed using video footage the same game previously observed L, in the presence of one researcher (with at least 8 weeks between observations and without any previous information about the game in analysis) that allowed them the possibility to stop the images, to rewind the video or to change the video projection velocity; (c) in video (V). V context implied that the observers analyzed the video of one

game, in the presence of one researcher, with the possibility to stop the images, to rewind the video or to change the video projection velocity.

The study was performed during the 2016-2017 competitive season, for games from the main football Portuguese competition, Liga NOS (1st Portuguese league). The video footage of each game was randomly selected. All the videos used for analysis were filmed in a superior plan, with an open angle of the game and maintaining the time-sequence of the game and the sound of the stadium. More mediatic games that were connected with controversial decisions of referees (penalties, offside game incidents, goals disallowed, goals originating from a foul incorrectly judged, red cards) according to the researchers' perception, were automatically excluded from the study in order to not promote bias in the RMOs assessment.

The referee's performance live and using video was registered in an official assessment sheet that was produced by the Portuguese (FA) for RMOs. All the registrations *live* and in *video* followed the recommendations of evaluation by Portuguese (FA) for RMOs.
The evaluation of the performance of referees was based on the analysis of key performance indicators (KPI) for each of the three main criteria: a) general evaluation; b) game difficulty and intervention; and c) disciplinary actions and game management. Each criterion was subdivided in different variables as described in Table 1.

Table 1 - Description of the analyzed KPI for each main criterion considered

Variables	Description	Evaluation
General evaluation		
Final Score	General performance of the referee	The score was between 10.00 (excellent performance) and <7.00 (very poor performance) according to UEFA rate
Game difficulty and intervention		
Difficulty	Difficulty and the incidences of the game registered for the 1st and 2nd half of the game	The score considered the critical incidents and the difficulty to manage the game. Easy (E = 1), Normal (N = 2), and Difficult (D = 3)
Personality	Referee personality during the game	The score considered the number of positive (P = 1) and negative (N = 2) interventions for game and players' management
Physical Condition and Position	Referee physical capacity	The score considered the number of negative situations registered and classify them in three groups <4=0; 4 =1; >4=2
Referee Teamwork	Referee and the assistant referees teamwork	The score considered the number of negative situations registered and classify them in three groups <3=0; 3 =1; >3=2

Disciplinary actions and game management		
Disciplinary actions	Disciplinary actions during the game that include the cards displayed by the referee were registered	The score considered the number of yellow and red cards showed by the referee.
Application of game laws P/N	Degree of application of the game laws	The score considered the number of positive (P = 1) and negative (N = 2) interventions registered related with the game law
Disciplinary control P/N	Disciplinary control of the referee during the game	The score considered the number of positive (P = 1) and negative (N = 2) interventions registered related with disciplinary control

(P): Positive; (N): Negative

1.1. Data analysis

We conducted an analysis of agreement of the evaluation by the RMOs (a) general evaluation; b) game difficulty and intervention; and c) disciplinary actions and game management) and compared L vs. LV, L vs. V as well as LV vs. V. Both Kolmogorov-Smirnov and Shapiro-Wilks Normality tests were used to verify the normality assumptions of each KPI. The KPI "Final score" violated the principles of normality. To obtain the normality of the data, the transformation was performed in two steps: a) calculation of mean and standard deviation (SD) of the values of the untransformed variables; b) calculation of the Intraclass Correlation Coefficient (ICC) and the Bland and Altman graph, based on the variables after the transformation, with the correspondent 95% confidence limits (CL). Thresholds for ICC were <0.19, poor; 0.20-0.29, weak; 0.30-0.59, moderate; 0.60-0.79, substantial; and >0.80, very strong (Lawrence & Lin, 1989).

Regarding the KPIs game difficulty and intervention (Difficulty, Personality, Physical Condition and Positioning, and Team work), Kappa de Fleiss' (K) was used with the correspondent 95% CL. Thresholds for K were 0-0.19, poor; 0.20-0.39, fair; 0.40-0.59, moderate; 0.60-0.79, substantial; and >0.80, almost perfect (Fleiss, Cohen & Everitt, 1969). With respect to the KPIs disciplinary actions and game management (Disciplinary actions; Application of game laws; Disciplinary control), Lin's correlation coefficient of agreement' (ρc) was calculated, with the correspondent 95% CL. Thresholds for ρc were <0.65 poor; 0.66-0.80 moderate; 0.81-0.90, substantial; and >0.90, almost perfect (Lawrence & Lin, 1989). Statistical analyses were performed using SPSS 24 software (IBM SPSS, USA).

Results

For all KPIs, the results revealed a higher agreement between observation L vs. LV than between L vs. V and particularly between LV vs. V. However, relevant differences on the magnitude of the agreement were observed between methods of observation for each KPI.

For the KPI final score, the results revealed a substantial agreement between L vs. LV and a moderate agreement between L vs. V and between LV vs. V. Greater mean values were noticed for L than for LV and V (L>LV>V) (see table 2).

Table 2 - Mean values of the criteria general evaluation for L, LV and V and Intra and Inter-observer analysis

Variable	L (Mean ± SD)	LV (Mean ± SD)	V (Mean ± SD)	ICC, CL (95%), Sig		
				L vs LV	L vs V	LV vs V
Final Score	8.18 ± 0.31	8.17 ± 0.35	8.13 ± 0.28	0.73, [0.51, 0.85], p<0.001	0.45, [0.15, 0.68], p<0.005	0.31, [-0.03, 0.58], p<0.05

Regarding the KPIs game difficulty and intervention, the results revealed higher agreement between L vs. LV than between L vs. V or LV vs. V for all the variables. However, different levels of agreement were observed between L and LV and between LV and V according to the variables observed. A decrease in the level of agreement on the variable difficulty was observed between first and second part. A substantial agreement between L vs. LV, and a moderate agreement between L vs. V and LV vs. V were observed for the variables Difficulty 1st part. A moderate agreement between L vs. LV, and a poor agreement between L vs. V and LV vs. V was observed for the variables Difficulty 2nd part. In the first part of the game higher values of difficulty were scored in V (V>LV>L), while in the second part the LV presented greater values and video the worst (LV>L>V) (see Table 3).

A substantial agreement between L vs. LV, a fair agreement between L vs. V and LV vs. V was observed for the variable Personality. Greater mean positive values were noticed for V than for LV and L (V>LV>L). A fair agreement between L vs. LV, and a poor agreement between L vs. V and LV vs. V was observed for the variable Physical Condition and Position. Greater mean values were noticed for L than for LV and V (L>LV>V). Interestingly, the low level of agreement for Personality and Physical Condition and Position was observed between L vs. V (see Table 3).

For the KPI referee teamwork, the results revealed a fair level of agreement between all conditions. Greater mean values for L than for LV and V (L>LV=V). The biggest changes on observations of teamwork were observed for the identification of negative

and neutral behaviours with the higher negative behaviours being recorded for V (V>L>LV) (see Table 3).

Table 3 - Results of the KPIs for difficulty and intervention for L, LV and V and correspondent comparisons

Variable	L			LV			V			K, CL (95%)		
	n (%)			n (%)			n (%)			L vs LV	L vs V	LV vs V
Difficulty 1st	E	N	D	E	N	D	E	N	D	**0.78, [0.44, 1.12]**	**0.52, [0.17,0.86]**	**0.43, [0.09, 0.76]**
	0	94.1	5.9	0	91.2	8.8	0	85.3	14.7			
Difficulty 2st	E	N	D	E	N	D	E	N	D	**0.57, [0.23, 0.90]**	**0.14, [-0.19,0.48]**	**0.11, [-0.23, 0.43]**
	0	79.4	20.6	0	76.5	23.5	0	82.7	17.6			
Personality	P	N	-	P	N	-	P	N	-	**0.74, [0.41, 1.07]**	**0.28, [-0.06,0.62]**	**0.33, [-0.01, 0.67]**
	76.5	23.5	-	79.4	20.6	-	82.4	17.6	-			
Physical Condition_Position	Neg	Neutral	Pos	Neg	Neutral	Pos	Neg	Neutral	Pos	**0.35, [0.15, 0.54]**	**0.02, [-0.18,0.22]**	**0.16, [-0.05, 0.38]**
	26.5	44.1	29.4	23.5	55.9	20.6	20.6	67.6	11.8			
Referee Teamwork	Neg	Neutral	Pos	Neg	Neutral	Pos	Neg	Neutral	Pos	**0.33, [0.08, 0.58]**	**0.28, [0.06, 0.50]**	**0.22, [-0.01, 0.45]**
	14.7	52.9	32.4	8.8	61.8	29.4	23.5	47.1	29.4			

Neg: Negative Pos: Positive
E: Easy: N: Normal; D: Difícult

Regarding the disciplinary actions and game management, the results revealed a higher agreement between L vs. LV compared to L vs. V and LV vs. V for all the variables (see table 4). However, the values of agreement across variables vary as a result of the different evaluation methods.

For the variable disciplinary actions, the results revealed a substantial agreement between L vs. LV, L vs. V as well as between LV vs. V. Higher mean values were recorded for L than for V and LV (L>V>LV) (see Table 4).

For the variable positive application of game laws, the results of the revealed a moderate agreement between L vs. LV and a poor agreement between LV vs. V, as well as L vs. V. Higher values were found for L than for LV and V (L>LV>V) (see Table 4).

For the variable negative application of game laws, the results revealed a moderate agreement between L vs. LV and a poor agreement between LV vs. V, as well as L vs. V. In opposition to previous results, greater mean negative behaviours higher negative values were recorded for V compared to L and LV (V>L>LV) (see Table 4).

For the KPIs positive disciplinary control, the results revealed a substantial agreement between L vs. LV and a poor agreement between LV vs. V, as well as between L vs. V. Higher values were registered for L and V than for LV (L=V>LV) (see Table 4).

Similarly, the KPI negative disciplinary control revealed a higher agreement between L vs. LV than between L vs. V as well as between LV vs. V. Greater mean negative behaviours were registered for V than for LV and L (V>LV>L) (see Table 4).

Table 4 - Mean values of the criteria disciplinary actions and game management for L, LV and V and correspondent comparisons.

Variable	L (Mean ± SD)	LV (Mean ± SD)	V (Mean ± SD)	ρ c, CL (95%)		
				L vs LV	L vs V	LV vs V
Disciplinary actions	6,97±2,04	6,47±2,33	6,56±1,99	**0.89, [0.79; 0.94]**	**0.86, [0.76; 0.93]**	**0.88, [0.78; 0.94]**
Application of game laws P	1,76±1,21	1,65±1,47	1,32±1,17	**0.78, [0.62; 0.88]**	**0.45, [0.16; 0.67]**	**0.43, [0.13; 0.66]**
Application of game laws N	0,88±1,09	0,76±0,85	1,12±0,95	**0.66, [0.43; 0.81]**	**0.04, [-0.28; 0.36]**	**0.10, [-0.21; 0.40]**
Disciplinary control and players management P	0,56±0,86	0,53±0,83	0,56±0,93	**0.81, [0.65; 0.90]**	**0.47, [0.16; 0.69]**	**0.51, [0.21; 0.72]**
Disciplinary control and players management N	0,88±0,77	1,03±1,11	1,12±0,91	**0.56, [0.31; 0.74]**	**0.22, [0.10; 0.51]**	**0.20, [0.13; 0.49]**

P: Positive; N: Negative

Discussion

The primary aim of this study was to examine the level of agreement between the most usual method of observation, live, and the other possible methods of observation live + video and using only video. The results indicated for all the variables the higher level of agreement between L vs. LV. The level of agreement between L vs. V and between LV vs. V varies according to the variable in analysis.

In line with our expectations, the use of different methods of observation to evaluate a referee's performance influences the judgement and the RMOs assessment (Halsey, 2007). From our results, particularly the combination of live and video evaluation seems to be the more reliable and balanced method of observation. The observation of the game live previous to the exposure to the video, even with at least 8 weeks of interval, seems to improve the agreement between analysis and to balance the final score of each variable (Araújo, Davids, Cordovil, Ribeiro & Fernandes, 2009). In fact, the observations associated with the LV for almost of the variables were generally positioned between L and V reference values. The combination of live and video footage

appears to allow RMOs to improve the perception of the events that convey information to better support the assessment of a referee's performance (Araújo et al., 2009).

Interestingly, and in line with previous research, it seems that the visual system can benefit from the actual or previous use of 3D information, especially in the case of complex and dynamic tasks such as football game (Put et al., 2014). That is, the judgement and decision-making of individuals are not only influenced by the characteristics of a given contexts of observation but also by the sequence of perceptual experiences used to access the available information. In line with this reasoning, the possibility to assess a referee performance *live* appears to facilitate subsequent judgments probably due to its richer conditions for active perception, such as adjustable 3D visual perspective of the observer perceiving beyond what can be captured by a video camera, identification and attunement to the game critical moments, social environment (public, players, coaches), off-ball information, etc.

The assessment of performance from expert observers incorporates different dimensions of a referee's performance (Mathers & Brodie, 2011; Pina et al., 2019; Russell et al., 2018). Each criterium is subdivided in different variables to allow for a complementary evaluation of different competencies of the referee throughout the game (Helsen & Bultynck, 2004). Accordingly, different variables require different information sources and even different perspectives of observation to avoid any bias (Dohmen & Sauermann, 2016; Gilis et al., 2008; Halsey, 2007).
The use of a LV method of observation could promote the combination of both perspectives and a careful technical analysis supported by the contextual environment experienced live. Considering the example of football referees (Pina et al., 2019), and the present results, we argue that the judgement and evaluation of RMOs needs to consider the specific context of the match. However, more research is needed to achieve a deeper comprehension of the impact of contextual variables on the decision-making process of RMOs.

Our results clearly revealed that, with exception of the variables difficulty and personality, L tend to rate all the other variables with higher values compared to V which strengthens the importance to prioritize L context of observation. In opposition, for the variable's negative application of game laws, negative disciplinary control and players management, the V revealed higher values. That is, it seems that the analysis live, in which RMOs have richer information about the game environment, tends to favour the overall score and the score of almost all the single variables. Despite of some agreement in analysis, the L scores are lower for the variable's difficulty and

personality. Such disagreement with the other methods may be explained by the use of more sources of information when the analysis of the game is L instead of in V (Halsey, 2007). Without the crucial game information and surrounding context, observers miss relevant information that may increase biases in referee's assessment.

In this way, it seems that the use of a unique perspective of analysis promotes bias in analysis, which is dependent of the context of observation. The combination of live with video analysis for the assessment of referees' performance, as is the case for most of the national and international leagues (Mathers & Brodie, 2011), could improve the RMOs reliability of analysis. One limitation of this study is that the combination of live with video analysis were interspersed by at least 8 weeks between observations. Further knowledge is required to understand the optimal interval of time between live with video analysis that may better improve the RMOs reliability of analysis.

In the Referee Convention guidelines of Europe's football governing body, UEFA, it is mentioned that match observations have to be organized *live* in the stadium. However, as from this season on, they introduced a new evaluation system that combines both live observation and video observation by two independent observers. Our results revealed that the combination of both L and V methods independently could promote higher levels of disagreement in the analysis, particularly in the variables personality, physical condition and negative application of game laws. Therefore, unfortunately, in the English Premier League, match observations are organized using video only. It should be clear that any match evaluation that disregards the live context, may result in an undesirable level of divergence that is not accurate and/or representative, particularly in variables that require richer conditions for active perception, identification and attunement to the game critical moments, social environment (public, players, coaches), off-ball information, etc…

Further research is required to understand the impact of the use of such independent methods in performance enhancement and career development of referees (Fruchart & Carton, 2012; Mathers & Brodie, 2011). As a limitation of this study we consider that it was not possible to identify the most accurate method of analysis to evaluate referees' performance. Further research is required indeed to improve the understanding of the accuracy of analysis for each of the various evaluation contexts. In this line of reasoning, the analysis of the assessment of RMOs could take advantage of analysing additional contextual variables such as game result, teams' aggressiveness, number of disciplinary actions, number of fouls or other candidate variables that may show a potential effect on the process of analysis of the performance of football referees.

Conclusion

In conclusion, this study reported different levels of agreement in the judgment and evaluation of the referee's performance between live, live + video and video only. The use of different contexts of observation may critically impact on the final score of a referee match performance. In particular, we found that the use of live context of observation is associated with a higher level of score for almost all of the variables except personality and perceived difficulty of the game. Conversely, the video only analysis seems to be associated with lower scores for match performances.

Our results suggest that the combination of live and video analysis allow RMOs to be more perceptually attuned to the events that convey relevant information to better support the assessment of referee's performance. This work offers evidence for the redesign of official observation procedures and policies of football governing bodies that are in charge in charge of referee's match evaluations.

Capítulo 4 – Identification of key predictors of Football referee's assessment: A case study in Portuguese league

Abstract

The purpose of this study was to evaluate the relationship between the assessment of football referees game performance and key performance indicators related with two main dimensions: i) game difficulty and game intervention and ii) disciplinary actions and game management. Thirty-four elite Referee Match Observers (RMO) from the Portuguese (FA) voluntarily participated in the study, and 34 RMO reports from the Portuguese Major League, season 2016-2017 were used for analysis. The data from the RMO reports correspondent to each game was categorized in two groups according to referee's assessment score: i) Referees with lower score; and ii) Referees with higher score. A multivariable binary logistic regression model was used to assess the relationship between the Key Performance Indicators (KPI) used by RMO to characterize game difficulty, game intervention, disciplinary actions, and game management in relation to game referee's assessment score. The model accounts for 60.4 % of the explanation of the game referee's assessment score. The results showed that the negative application of game laws and referee teamwork were the KPI that most contributed for the definition of the game referee's assessment score. The results did not reveal any differences for the variables game difficulty, physical condition, position, disciplinary actions, positive application of game laws and positive disciplinary control and players management. In conclusion, the variables evaluated by the RMO have different weightings for the assessment of referee's game performance. Interestingly, the variables that most discriminate game referee's assessment score were related with a negative evaluation of disciplinary criteria and teamwork.

Keywords: referee evaluation; referee teamwork; performance analysis; decision making; key performance indicators

---

Carvalho, V., Esteves, P.T., Nunes, C., Costa, P., Helsen, W.F., Travassos, B. (Under revision). Identification of key predictors of Football referee's assessment: A case study in Portuguese league

Introduction

In football, the referee and assistant referee aim to apply the laws of the game according to game demands (Catteeuw et al., 2009). In fact, football referees do not only need to make accurate decisions, but also need to apply proper disciplinary sanction on players according to the laws of the game (Catteeuw et al., 2009; Elsworthy et al., 2014), and to preserve the integrity of the game (Mascarenhas et al., 2002). Helsen and Bultynck (2004) observed that referees have to make approximately 200 to 250 decisions per match. From the total of decisions, 28% are related with the Law 12 (Fouls and misconduct), and about 31% with the Law 15 (Throw-in). Almost all of these decisions are made in dynamic environments, in a short period of time, with reduced visual information available and in, sometimes, with the need to depict players' deceptive movements (Lex et al., 2015). More than that, refereeing not implies only the judgement and decision-making but also the management of the interpersonal relations with the players and other football agents to guarantee a good management environment over the game (Mallo et al., 2012). Thus, refereeing performance is highly dependent on levels of expertise to appropriately the manage game (Gilis et al., 2008). Thus, the number of years of experience of the referees as well as the hours of practice per week, the number of matches or even the difficulty and the context of such matches are clearly important factors that may impact on referee performance (Helsen et al., 1998).

Previous research have revealed that expert referees develop effective anticipatory strategies to improve the accuracy and the time of their decision making (Mather, 2008), show a better physical condition, a lower level of stress and make a lesser number of calls on penalties than lower-level referees. In fact, the best referees, being more competent than their counterparts not only in terms of their decision-making, but also in their better position on the field according to the dynamics of the game (MacMahon, Starkes & Deakin, 2007), allowing a better perception of players information, the ball and the assistant referees (Ghasemi et al., 2009). Aoba, Yoshimura, Miyamori & Suzuki (2011) also reported that more experienced referees cooperate more with assistant referees than less experienced referees, reducing the distance and the grey zones between them, and involving them more in the decisions. In this line of reasoning Helsen and Bultynck (2004) indicated that 64% of all decisions of the best referees are based on teamwork.

Furthermore, the elite referees tend to express a greater capacity to better deal with the characteristics and requirements of each game, and to the contextual factors that

characterize it (Carvalho, Esteves, Nunes, Mendez & Travassos, 2020). Consequently, players, coaches and other agents reported high levels of satisfaction with the performance of elite referees (Guillén & Feltz, 2011), in relation to their strong personality and ability of game management, with consequential improvements on the final performance (Bar-Eli, Plessner & Raab, 2011).

In order to proceed with the evaluation of the football referee, the Referee Convention guidelines of UEFA mentioned that the assessment of referee game performance should be accomplished by a Referee Match Observer (RMO) by means of *live* observation in the stadium. The RMO should observe the match and register in an official assessment sheet the referee's assessment score according to two main criteria of analysis - game difficulty and game intervention; and disciplinary actions and game management - in which specific key performance indicators (KPI) should also be registered (Mathers & Brodie, 2011). The analysis comprises the evaluation of the KPIs that allow a complementary evaluation of different competencies of the referee throughout the game (Helsen & Bultynck, 2004).

Despite of this process being used for a large number of years in almost of Football Associations, it is an unknown relationship between the game referee's assessment score and the scores of the different KPIs. Due to the impact of the game referee's assessment score atributed to the referee at the end of each game in their upgrade, downgrade or maintenance in a certain level (Fruchart & Carton, 2012), further research is required regarding the process of assessment of football referees. Particularly, there is a need to further understand the relationship and the weight of the KPIs on game referee's assessment score. Thus, the purpose of this study was to evaluate the relationship between the best and worst game referee's assessment score and the KPIs evaluated and registered on the RMO official assessment sheet. It was expected to identify different the weights of each KPI evaluated for different game referee's assessment scores.

Methods

Participants

Thirty-four elite Referee Match Observers (RMO) from the Portuguese (FA) voluntarily participated in the study (Age – 52.62 ± 8.95; Number of years as observer – 12.11 ± 6.81; Number of years as observer in 1st league – 7.18 ± 5.69). An informed and written consent was provided to the Portuguese (FA) Referee Council, as well as to all the participants, before the beginning of the study. All participants were informed about

the goal of the study and notified that they could withdraw from the study at any time. The study protocol followed the guidelines stated in the Declaration of Helsinki and was approved by the Local Ethics Committee

Data collection

Thirty-four games from the Portuguese main league 2016-2017 were analyzed in the stadium by an expert RMO and submitted to the Portuguese (FA) in 24 hours after the end of the game. The RMO registered the analysis in an official assessment sheet that was produced by the Portuguese (FA). The evaluation of the performance of referees was divided in a final game referee's assessment score between 0-10, and the evaluation of two main dimensions (game difficulty and game intervention; disciplinary actions and game management) in which specific KPIs were registered (see Table 5).

Table 5 - Description of the analyzed KPI for each main criterion considered.

Variables	Description	Evaluation
General evaluation		
Final Score	General performance of the referee	The score was between 10.00 (excellent performance) and <7.00 (very poor performance) according to UEFA rate
Game difficulty and intervention		
Difficulty	Difficulty and the incidences of the game registered for the 1st and 2nd half of the game	The score considered the critical incidents and the difficulty to manage the game. Easy (E = 1), Normal (N = 2), and Difficult (D = 3)
Personality	Referee personality during the game	The score considered the number of positive (P = 1) and negative (N = 2) interventions for game and players' management
Physical Condition and Position	Referee physical capacity	The score considered the number of negative situations registered and classify them in three groups <4=0; 4 =1; >4=2
Referee Teamwork	Referee and the assistant referees teamwork	The score considered the number of negative situations registered and classify them in three groups <3=0; 3 =1; >3=2
Disciplinary actions and game management		
Disciplinary actions	Disciplinary actions during the game that include the cards displayed by the referee were registered	The score considered the number of yellow and red cards showed by the referee.
Application of game laws P/N	Degree of application of the game laws	The score considered the number of positive (P = 1) and negative (N = 2) interventions registered related with the game law
Disciplinary control P/N	Disciplinary control of the referee during the game	The score considered the number of positive (P = 1) and negative (N = 2) interventions registered related with disciplinary control

(P): Positive; (N): Negative

Data analysis

The data from the official assessment sheet produced by each game was categorized in final game referee's assessment score and according to the KPI registered for each two main criteria (game difficulty and game intervention; disciplinary actions and game management). KPIs were identified as independent variables while the final game referee's assessment score was identified as the dependent variable. A preliminary analysis revealed no missing values neither outliers. Moreover, for statistical purposes, a dummy variable of the game referee's assessment score was created and the two groups were generated: Lower Final Score (LFS) - games with a final score lower than the median Final score, n = 19; Higher Final Score (HFS) - games with a Final score higher than the median Final score, n = 15.

Subsequently, a descriptive analysis of the KPIs according to the final game referee's assessment score was performed (continuous variables were described by means and standard deviation and categorical variables by absolute and relative frequencies). A multivariable binary logistic regression model was used to assess the relationship between the KPIs and the final game referee's assessment score. Only the variables whose univariate test was significant (p<0.05) were selected to the model. The nominal KPI were analyzed through the Fisher's exact test (since the assumptions underlying the use of the Chi-square test were not verified) and the continuous KPI by the *Mann-Whitney U test* (since the normality assumption underlying the use of the independent samples *t-test* was not verified). Normality was checked using the Shapiro-Wilk test.

The Forward stepwise (Likelihood Ratio) selection method was considered, and the results were reported by odds ratio (OR) estimates and their 95% confidence intervals (CI). To evaluate the model's power of explanation the *Nagelkerke's* R^2 was used. Its interpretation as an effect size measure was made based on the following criteria: 0.02–0.13 small, 0.13–0.26 medium, and >0.26 large effect size. The model's goodness of fit was assessed through the Hosmer-Lemeshow test and the Area Under the Curve (AUC) through Receiver Operating Characteristic (ROC) was used to evaluate the discriminant capacity of the model. The ROC curve was performed using the predicted probabilities of each variable. All the statistical analyses were performed using SPSS 24.0 and statistical significance for rejecting the null hypothesis was set at p<0.05.

Results

Regarding the analysis of the dimension game difficulty and game intervention, results revealed significant differences only between the KPIs personality and referee teamwork according to the final game referee's assessment score ($p<0.05$). The HFS revealed higher number of positive interventions regarding personality and higher positive actions of referee teamwork than LFS (see Table 6).

Table 6 - Results of the KPI for difficulty and intervention for LFS and HFS.

Variable	LFS n (%)			HFS n (%)			*p-value#1*
Difficulty 1st	E	N	D	E	N	D	**1**
	0	94.7	5.3	0	93.3	6.7	
Difficulty 2st	E	N	D	E	N	D	**0.672**
	0	84.2	15.8	0	73.3	26.7	
Personality	P	N	-	P	N	-	**0.005***
	57.9	42.1	-	100	0	-	
Physical Condition_Position	Neg	Neutral	Pos	Neg	Neutral	Pos	**0.413**
	26.3	52.6	21.1	26.7	33.3	40	
Referee Teamwork	Neg	Neutral	Pos	Neg	Neutral	Pos	**0.007***
	21.1	68.4	10.5	6.7	33.3	60	

#1 – Exact Fisher test; * $p<0.01$; ϕ Cramers' V

The analysis of the dimension disciplinary actions and game management revealed significant differences only for KPIs negative application of game laws and negative disciplinary control and players management variables and the final game referee's assessment score ($p<0.05$). The HFS revealed lower number of negative application of game laws and also lower number of negative disciplinary control and players management (see Table 7).

Table 7 - Results of the KPI for criteria disciplinary actions and game management for LFS and HFS.

Variable	LFS (Mean ± SD)	HFS (Mean ± SD)	*p-value#2*
Disciplinary actions	7±2.03	6.93±2.12	**0.876**
Application of game laws P	1.747±1.22	2.13±1,13	**0.127**
Application of game laws N	1.37±1,21	0.27±0.46	**0.002***
Disciplinary control and players management P	0.53±0,91	0.60±0.83	**0.598**
Disciplinary control and players management N	1.16±0.83	0.53±0.52	**0.021****

#2 –U Mann-Whitney test; * $p<0.01$; ** $p<0.05$

In the multivariable binary logistic regression model, the independent KPIs personality and negative disciplinary control and players management were excluded from the model ($p>0.05$). The overall model presents a well-fitting value ($p_{Hosmer\text{-}Lemeshow}>0.05$), a good correct global classification (79.4%) and its discriminant capacity is also quite good with the AUC ranging between 0.788 – 0.998, with a 95% confidence level. The model accounts for 60.4 % of the explanation of the game referee's assessment score (*Negelkerke* R^2=0.604), corresponding to a large effect size (see Table 8).

The results revealed that for each negative application of game laws, the chances to achieve a HFS decreased by 85% (*OR* = 0.15, $CI_{95\%}$= [0.029–0.75]). Also, the analysis of the referee teamwork revealed that a referee with a neutral evaluation has 3 times more chances to achieve a HFS than a referee with a negative evaluation (*OR* = 3.35, $CI_{95\%}$ = [0.22–50.68]). A referee with a positive evaluation has 46 times more chances to achieve a HFS than a referee with a negative evaluation (*OR* = 45.84, $CI_{95\%}$ = [1.67-1258.35]) (see Table 8).

Table 8 - Estimated regression coefficients from binary logistic regression analysis, only for significant variables (final score as dependent variable).

Variables	*B*	OR (CI 95%)	p	AUC (CI 95%)
Application of game laws N	-1.92	0.15 (0.029-0.750)	.021	
Referee Teamwork				
Negative[b]		1		.893 (.788 - .998)
Neutral	1.21	3.35 (0.22-50.68)	.383	
Positive	3.83	45.84 (1.67-1258.35)	.024	
Nagelkerke R^2 =0.604	Hosmer-Lemeshow test (p=.961)		Correct global classification= 79.4%	

Note. OR - Odds ratio; [b] Reference category; AUC - Area under the ROC curve

Discussion

The purpose of this study was to evaluate the relationship between game referee's assessment score (High final score vs Low final score) and the KPIs evaluated and registered on the RMO official assessment sheet associated to two different dimensions: i) game difficulty and intervention ii) disciplinary actions and game management.

Overall, the comparison between LFS and HFS revealed differences in personality, referee teamwork, negative application of game laws, negative disciplinary control and

player management KPIs. The results from multivariable binary logistic regression model revealed that the KPIs negative application of game laws and referee teamwork were the ones that significantly defined the final game referee's assessment score. The model accounts for 60.4 % of the explanation of the game referee's assessment score.

Characterization of game LFS and HFS referee's assessment

The development of referee excellence and skills are strongly dominated by the existence of evaluation/rating processes, being of central preponderance for their upgrade, downgrade or maintenance in a certain competitive level (Fruchart & Carton, 2012). Therefore, in this study we aimed to identify the KPIs that best discriminate LFS from HFS groups based on the KPIs from the RMO official assessment sheet. It was observed that the HFS revealed lower errors in the application of game laws and in the application of disciplinary control and players management than LFS. The HFS referees also revealed higher number of positive interventions regarding personality and referee teamwork than LFS.

Based on these results it is possible to sustain previous assumptions that the best referees are the ones that demonstrate greater ability to correctly apply the disciplinary criteria described in the laws of the game (Helsen et al., 1998). Our results clearly demonstrate that the HFS show better performance in terms of the personality demonstrated in game intervention, as well as in the teamwork level, in comparison with the LFS. In fact, Collina (2003) considered that the best referees are those who have the greatest knowledge and skills related to the regulation and disciplinary aspects of the game, but also to the ability to manage the game according to its dynamics and the contextual factors that interfere with it (Carvalho et al., 2020). Thus, the elite referees that revealed better performance seem to be able, probably due to a better capacity for cooperation with the assistant referees, to judge and decide with higher accuracy and precision (Aoba et al., 2011).

In opposition, the results of the analysis of the disciplinary criteria and the management of the game revealed that the LFS referees tended to present more errors (negative) in terms of decision making, in the application of the game laws, and in the control disciplinary and management players. Each decision and intervention of the referee in such critical moments during the game (penalty, controversial offside) can dramatically affect the result of the match (Halsey, 2007). Thus, particularly the negative application of the game laws and the negative control disciplinary and management players should be avoided to improve referees' performance.

Key performance indicator values of referees' performance according to game referee's assessment score

The game referee's assessment from RMO incorporates different dimensions of a referee's performance (Pina et al., 2019; Russell et al., 2018). In general, the analysis comprises the definition of the final score of referee assessment but also the evaluation of the KPIs such as the technical and disciplinary decisions, the difficulty of the game, the positioning, as well as the capacity of game management and collaboration with assistant referees (Mathers & Brodie, 2011). Despite of this process being used for a large number of years in almost of Football Association, the relation between the final score of referee assessment and the scores of the different criteria evaluated remained somewhat unclear. The results of this study build on this lack of knowledge to reveal that the KPIs negative application of game laws and referee teamwork explained 60.4 % of the final game referee's assessment score. Mathers and Brodie (2011), proposed that the performance of elite referees implicate a multifaceted integration of several distinct processes that depends on a proper level of interaction with assistant referees. Therefore, and instead of trying to perceive all the occurring incidents in the field, sharing information and decisions, in the context of an elite refereeing teamwork, is essential to improve referees' performance (Catteeuw et al., 2009). The results of the applied model revealed that a referee who shows a negative score on the KPI teamwork, has 3 times less chance to be in the HFS group, than a referee who scores neutral on teamwork. Likewise, a referee who gets a negative score on the KPI teamwork has 46 times less chance to be in the HFS group, compared to the referee with a positive score.

As mentioned previously, football referees are not only required to make accurate decisions and disciplinary sanction (Elsworthy et al., 2014) based on the laws of the game, but also to ensure an appropriate management of the game (Mascarenhas et al., 2002). In other words, during a game, a referee who is being negatively evaluated in the application of the game laws, more specifically in the disciplinary action and game management, has 85 times less change to be in the HFS group, compared to the referee with a positive score. In fact, the best referees were characterized by critically deciding within a time window and supporting the decision on relevant information about the players, the ball and the fellow assistant referees (Ghasemi et al., 2009; Helsen & Bultynck, 2004). Further research should be developed to understand if not only the number, but also the nature of disciplinary actions may influence the game referee's assessment score.

Despite the fact that this study stands as the first attempt to gain a further understanding on the assessment of game referee's performance according to the contribution of each category for the computation of the final score, there are some limitations that should be declared. First, data collection occurred in a single country and during only one season so, in a certain way, it is representative of a quite specific context. Also, the relationship between the KPIs analyzed and the disciplinary and game management events such as, for example, a penalty kick or a display of a correct or incorrect red card were not considered and, in this way, should be considered in future studies.

Conclusion

In conclusion, this study suggests that the variables evaluated by the RMO have different weightings for the calculation of game referee's assessment score. The model accounts for 60.4 % of the explanation of the game referee's assessment score based on the variables teamwork and the negative application of the game laws.

However, the referees with the best game referee's assessment score still revealed higher number of positive interventions regarding personality and lower number of negative disciplinary control and players' management compared to the referees with the worst game referee's assessment score. The results did not reveal any differences for the variables game difficulty, physical condition_position, disciplinary actions, positive application of game laws and positive disciplinary control and players management. Based on this study, relevant stakeholders such as Football Association can take advantage of the results here found to improve the methods of assessment of referee game performance.

Capítulo 5 – A influência de variáveis contextuais e situacionais na classificação de árbitros de futebol de elite

Resumo

Este trabalho teve como objetivo avaliar a relação entre a classificação dos árbitros de futebol de elite em Portugal e as variáveis contextuais e situacionais dos jogos arbitrados no decorrer da época desportiva 2016-2017. No sentido de analisar qual a relação entre a classificação final e o nível da competição, caraterização do jogo, resultado do jogo e número total de cartões exibidos por jogo foi realizada uma regressão ordinal com função *Link* Logit. Os resultados, revelam que as variáveis contextuais, apresentam um efeito significativo sobre a classificação final, não se verificando efeitos significativos das variáveis situacionais, sobre a classificação final dos árbitros de futebol no final da época desportiva. Na globalidade o modelo revelou-se estatisticamente significativo (X^2_{LP} (5) = 40.299, p<0.001). A probabilidade de obtenção de melhor classificação final dos árbitros aumenta 54.2% com o aumento do número de jogos realizados na I Liga (*OR*=1.542) e aumenta 24.8% com aumento do número de jogos equilibrados (*OR*=1.248). Diminui 61.2% com jogos realizados sem equipas Top 3 (*OR*=0.388). Por último, em relação ao número total de cartões exibidos num jogo não se verificam efeitos significativos no ranking de desempenho dos árbitros. Em suma, os resultados reforçam a influência significativa que os fatores contextuais têm sobre a classificação e avaliação de um árbitro no final da época desportiva.

Palavras-chave: futebol; arbitragem; desempenho; avaliação; classificação

---

Carvalho, V., Esteves, P. T., Nunes, C., Mendez, C., & Travassos, B. (2021). A influência de variáveis contextuais e situacionais na classificação de árbitros de futebol de elite. Cuadernos de Psicología del Deporte, *21*(1), 216–224.
https://doi.org/10.6018/cpd.435081

Introdução

Fruto da grande variabilidade e dinâmica situacional que caraterizam o jogo de futebol, o árbitro de futebol de elite necessita realizar de forma continua decisões que lhe permitam, no decorrer do jogo, assegurar o cumprimento das leis do jogo (Elsworthy et al., 2014). Ao longo dos últimos anos, muitos estudos avaliaram a prestação do árbitro de futebol tendo por base a avaliação da tomada de decisão do árbitro (Elsworthy et al., 2014). Esta foi aliás, durante largos anos considerada a competência mais importante do árbitro (Helsen & Bultynck, 2004; Morris & O'Connor, 2016), sendo avaliada como um meio eficaz para diferenciar peritos e não peritos (Berry et al., 2008; Williams & Ericsson, 2005), pois, considera-se que reflete o nível de adaptabilidade do árbitro à variabilidade do próprio jogo (Urra et al., 2018; Kittela et al., 2019).

No entanto, esta tarefa não tem apenas por base o cumprimento estrito das leis do jogo (Guillén & Jiménez, 2001), sendo influenciada pela avaliação constante dos seus pares e do público em geral (Collina, 2003), necessitando o árbitro para isso de grande resiliência e motivação para o desempenho das suas funções (Folkesson, Nyberg, Archer & Norlander, 2002; Wolfson & Neave, 2007). A este propósito, Pina et al., (2019) consideram que o desenvolvimento de excelência e das competências do árbitro decorrem do desenvolvimento de três dimensões distintas, sendo altamente influenciadas pelo processo que as medeia, nomeadamente as classificações, promoções e as despromoções de categoria (Fruchart & Carton, 2012), o apoio psicológico, a pressão, a humilhação pública na televisão e nos jornais diários a que estão sujeitos (Weaver, 2001; Wilson, 2000), bem como os programas de formação e de treino existentes.

Mascarenhas et al., (2005), identificam cinco fatores que sustentam a excelência no desempenho da arbitragem: a) competências de personalidade e na gestão do jogo, b) julgamento contextual, c) conhecimento e aplicação da lei, d) características psicológicas de excelência e e) aptidão física, posicionamento e qualidades mecânicas. Estes fatores foram corroborados em outros estudos realçando que a experiência na arbitragem em contextos adversos (Catteeuw et al., 2009), competências na gestão do jogo (Sabatini, 2002) bem como competências de comunicação, de gestão das emoções (Dosseville, Laborde & Bernier, 2014), de inteligência emocional, e liderança para além do conhecimento de regulamentos e aspetos técnicos são fundamentais para a excelência na arbitragem (Sülün, 2013). Em suma, a prestação do árbitro de futebol é influenciada por fatores: a) contextuais que marcam o contexto de jogo e o ambiente que o rodeia; b) situacionais que ocorrem durante o jogo, fruto da dinâmica do mesmo

e da relação estabelecida entre árbitro, jogadores, treinadores e público; e c) individuais fruto da preparação e capacidade do próprio árbitro para gerir o jogo.

A respeito dos fatores contextuais que influenciam a prestação do árbitro de futebol, Nevill et al., (2002), revelam que, num estudo experimental, os árbitros que ouviam o barulho do público num clip de vídeo, eram influenciados na decisão sobre possível situação de infração da equipa da casa, do que os árbitros que assistiram aos clipes sem som ambiente. No entanto, embora fatores contextuais (equipa da casa ou público) pareçam influenciar a tomada de decisão do árbitro em competições nacionais (Buraimo et al., 2010; Buraimo et al., 2012), o mesmo não se verifica em competições internacionais com árbitros de elite. De modo semelhante, parece que o nível competitivo das equipas parece influenciar a gestão e preparação do próprio jogo (Tenreiro-Gavela et al., 2016), bem como o modo como os árbitros decidem a atribuição de um penalti (Erikstad & Johansen, 2020). Por outro lado, Jones et al., (2002) verificam que o conhecimento prévio dos árbitros acerca da agressividade das equipas influencia a exibição de cartões amarelos, isto é, à equipa considerada previamente mais agressiva eram mostrados mais cartões amarelos do que à equipa contrária.

No que diz respeito aos fatores situacionais, o modo como o árbitro realiza a gestão do jogo pode levar a que a mesma tomada de decisão ou uma ação disciplinar possa ser interpretada como um erro de julgamento face às leis de jogo, ou pode ser vista como uma decisão correta de modo a salvaguardar a gestão e fluidez do jogo (Askins, 2001; Grunska, 2001). Por exemplo, Brand, Schmidt e Schneeloch (2006) verificam num estudo experimental que, quando os árbitros viram uma sequência de clipes com infrações pela ordem em que ocorreram durante o jogo, assumiram sanções menos rigorosas e mais diversas, do que ao ver as mesmas como uma infração isolada. Plessner e Betsch (2001), revelam que a decisão de assinalar um penalti, poderá ser influenciada por decisões anteriores em situações similares, já tendo sido demonstrado em outros contextos desportivos que as “regras não escritas” podem influenciar as decisões dos árbitros (Plessner & Raab, 1999). De forma semelhante, Lex et al., (2015) revelam ainda que a ação disciplinar do árbitro através da exibição de cartão amarelo era mais vezes registada perante a vocalização do jogador e pela necessidade de gestão do jogo, do que propriamente fruto de uma infração às leis de jogo.
Fruchart e Carton (2012), em relação às caraterísticas individuais, indicam que os árbitros de nível inferior tendem a cometer mais erros e de alguma forma premeditada, tornando a gestão do mesmo mais difícil. No entanto, tal como referido anteriormente, esta parece ser uma estratégia de valorização individual fruto das caraterísticas que

assumem nomeadamente as classificações que dão origem às promoções e despromoções dos árbitros no final da época desportiva.

A respeito da avaliação da prestação do árbitro de futebol e consequentemente da classificação no final da época desportiva, Weaver (2001) considera que este processo influencia bastante a prestação do árbitro no decorrer do jogo, não o permitindo em algumas circunstâncias de adotar uma gestão do jogo mais eficaz e de acordo com o contexto em que este ocorre, em detrimento do estrito cumprimento das leis de jogo. As avaliações do desempenho do árbitro, são normalmente efetuadas por ex-árbitros experientes, que de acordo com as diretrizes da FIFA, UEFA e do próprio conselho de arbitragem da federação nacional avaliam e classificam o desempenho do árbitro de futebol, num determinado jogo. Esta avaliação é realizada tendo por base uma observação ao vivo ou em vídeo tendo por base grelhas de observação que consideram as seguintes categorias de observação: i) dificuldade e intervenção no jogo, no qual se tem em consideração fatores contextuais, capacidade individual e capacidade de posicionamento e colaboração entre membros da equipa de arbitragem; ii) ações disciplinares e gestão do jogo, no qual se consideram aspetos relacionados com ação disciplinar, aplicação das leis de jogo e controlo disciplinar. Sabendo da pertinência e influência destes processos de observação na prestação imediata dos árbitros, bem como no desenvolvimento da sua carreira desportiva, nada se sabe, no entanto, sobre a possível influência de algumas condições contextuais e situacionais na avaliação do desempenho do árbitro e consequentemente da sua classificação no final da época desportiva. De facto, não é conhecido até ao momento nenhum estudo que tenha verificado a influência de fatores contextuais ou situacionais na classificação obtida no jogo, bem como na respetiva classificação final dos árbitros no final de cada época desportiva.

Deste modo, este estudo teve como objetivo, analisar a relação entre a classificação dos árbitros de futebol de elite em Portugal, e as variáveis contextuais e situacionais dos jogos arbitrados, no decorrer da época desportiva 2016-2017. Como hipóteses consideramos que: i) a classificação dos árbitros de futebol é influenciada por variáveis contextuais (Sülün, 2013); os árbitros melhor classificados tendem a arbitrar jogos de competições de nível superior, entre as melhores equipas, e jogos mais equilibrados em comparação com árbitros pior classificados; ii) a classificação dos árbitros de futebol é influenciada pelo número de ações disciplinares no decorrer do jogo (Fruchart & Carton, 2012). Os árbitros melhor classificados tendem a apresentar menor número de ações disciplinares em comparação com árbitros pior classificados.

Métodos

Desenho

Este estudo teve como base um desenho empírico de natureza associativa que envolveu a comparação de variáveis experimentais entre diferentes grupos com o intuito de explorar associações de funcionalidade e eventuais diferenças (Ato, López & Benavente, 2013).

Participantes

Vinte e três árbitros de futebol (média de idades = 36,6 ± 4,08 anos) que constituíam o quadro de árbitros da 1ª categoria em Portugal na época desportiva 2016-2017, foram considerados para o estudo. Os últimos quatro classificados desciam de categoria. A cada árbitro foi atribuído um Código e registados quinhentos e cinquenta e oito jogos no âmbito da I Liga e II Liga. A cada árbitro foi ainda atribuída, a respetiva classificação final oficial da época desportiva, tendo por base o somatório da avaliação de cada jogo pelo Conselho de Arbitragem da Federação Portuguesa de Futebol. Este trabalho foi aprovado pelo comité de ética da Universidade da Beira Interior de acordo com a declaração de Helsínquia (2013).

Recolha de dados

Todos os dados foram selecionados e recolhidos, da base de dados digital privada, zerozero (www.zerozero.pt) e registados numa matriz de excel. Como variável dependente, foi considerada a variável classificação final dos árbitros, na época desportiva 2016-2017, através da divisão da classificação em tercis: 1º tercil (G1) – classificação superior (do 1º ao 8º classificado inclusive); 2º tercil (G2) – classificação intermédia (do 9º ao 16º classificado ambos inclusive); 3º tercil (G3) – classificação inferior (do 17º ao 23º classificado ambos inclusive).

Como variáveis independentes foram definidas variáveis contextuais: i) nível da competição (1 - I Liga; 2 - II Liga), ii) Caraterização nível competitivo do jogo (Outros x Outros: jogos sem equipas do Top 3 do campeonato; Top 3 x Outros: jogos com uma equipa Top 3 e outra equipa que não Top 3 do campeonato; Top 3 x Top 3: jogos entre equipas do Top 3 do campeonato) (a caraterização do nível competitivo do jogo foi considerada em três níveis, tendo por base os critérios de avaliação de desempenho do

árbitro de futebol em vigor: jogo com participação de pelo menos uma equipa Top 3 é considerado como um jogo de dificuldade mais elevado; jogo com participação de duas equipas Top 3 é considerado jogo de dificuldade superior; jogo sem participação de equipas Top 3 é considerado como um jogo de dificuldade normal); iii) resultado do jogo (1 – equilibrado - quando existe uma diferença até dois golos; 2 - empate, 3 – desequilibrado – quando existe uma diferença superior a dois golos); e variáveis situacionais relacionadas com a ação disciplinar do árbitro no decorrer do jogo: iv) número total de cartões exibidos por jogo (total de cartões amarelos e vermelhos exibidos por jogo).

Tratamento de dados

No sentido de analisar qual a relação entre a classificação final e o nível da competição, caraterização do nível competitivo do jogo, resultado do jogo e número total de cartões exibidos por jogo, foi realizada uma regressão ordinal com função *Link* Logit. A seleção da função *Link* foi baseada na frequência de distribuição das classes das variáveis dependentes (Marôco, 2014). O pressuposto do modelo da homogeneidade de declives foi validado (X^2_{LP} (5) =3.356, p=0.645). Os *OR* foram estimados para as classificações superiores, em comparação com as classificações inferiores. A análise estatística foi efetuada com recurso ao SPSS v 23.0 para MAC (IBM Corp.). Foi considerada uma significância estatística de 5%.

Resultados

Através da análise da Tabela 9, verificámos que, em geral, para todas as variáveis contextuais, os valores de G1 são mais elevados em comparação com o G2 e G3. De forma mais especifica, os árbitros G1 realizaram mais jogos de nível de competição superior (I Liga), realizaram mais jogos com equipas Top 3 e apresentaram valores médios mais elevados em termos de resultado de jogo quando comparados com o G2 e G3. Em relação às variáveis situacionais, os árbitros do G1 foram os que apresentaram menor número de cartões exibidos por jogo. Os resultados alcançados pelo G2, estão tendencialmente na mediana, entre o G1 e G3 (Tabela 9).

Tabela 9 - Variáveis contextuais e situacionais que influenciam a classificação final dos árbitros de futebol.

Variáveis		G3	G2	G1	Estimativa	*Wald* (p)	*OR* (IC 95%)
Variáveis contextuais							
Nível da competição	I Liga	42,14%	52,38%	63,33%	0.432	6.395 (0.011***)	1.542 ([1.102;2.154])
	II Liga	57,86%	47,62%	36,67%	0.000	-	Ref
Caracterização nível competitivo do jogo	Outros	93,71%	83,07%	73,81%	-0.947	15.269 (0.000*)	0.388 ([0.241; 0.624])
	Top 3	6,29%	16,93%	26,19%	0.000	-	Ref
Resultado do jogo	Equilibrado	67,92%	61,38%	56,67%	0.221	0.636 (0.425)	1.248 ([0.724; 2.148])
	Empate	21,38%	29,10%	32,86%	0.710	5.518 (0.019***)	2.033 ([1.125; 2.148])
	Desequilibrado	10,69%	9,52%	10,48%	0.000	-	Ref
Variáveis situacionais							
Número total de cartões exibidos por jogo		37,76%	33,71%	28,53%	0.020	0.329 (0.566)	1.020 ([0.952; 1.094])

Ref – Categoria de Referência; *p<0.001; **p<0.01; ***p<0.05

Os resultados demonstram que, na globalidade, o modelo se revelou estatisticamente significativo (X^2_{LP} (5) = 40.299, p<0.001). Os jogos realizados na I Liga aumentam a chance de obtenção de melhor classificação final dos árbitros em 54.2% quando comparados com os jogos realizados na II Liga (p=0.011; *OR*=1.542; IC 95%=[1.102;2.154]). Os jogos realizados sem equipas Top 3 diminuem em 61.2% a possibilidade dos árbitros obterem melhor classificação quando comparados com os jogos realizados com equipas Top 3 (p<0.001; *OR*=0.388; IC 95%=[0.241; 0.624]). Um resultado equilibrado do jogo aumenta em 24.8% a possibilidade de obtenção de uma melhor classificação, apesar de não significativa, (p=0.425; *OR*=1.248; IC 95%=[0.724; 2.148]), enquanto que um empate duplica a chance de uma melhor classificação final quando comparado com um resultado desequilibrado do jogo (p=0.019; *OR*=2.033; IC 95%=[1.125; 2.148]). Por último, em relação ao número total de cartões exibidos num jogo, não se verificam efeitos significativos no ranking de desempenho dos árbitros (p=0.566; *OR*=1.020; IC 95%=[0.952; 1.094]).

Discussão

O principal objetivo deste estudo, foi avaliar qual a relação entre as variáveis contextuais e situacionais em estudo, e a classificação final dos árbitros de futebol de elite em Portugal, no decorrer da época desportiva de 2016-17. De acordo com as hipóteses consideradas, os resultados evidenciaram, que a combinação entre as

variáveis contextuais e situacionais exploradas, por si só, explicam cerca de 40% da classificação final dos árbitros de futebol de elite. Especificamente, as variáveis contextuais, apresentam um efeito significativo sobre a classificação final dos árbitros de futebol, no final da época desportiva. Ao invés do esperado, não se verificam efeitos significativos das variáveis situacionais, número de ações disciplinares, sobre a classificação final dos árbitros de futebol no final da época desportiva.

Qualquer processo de avaliação de desempenho, e em particular na arbitragem, fruto do peso e influência que tem sobre os próprios árbitros (Pina et al., 2019), deve ter como objetivo a avaliação dos fatores que sustentam a excelência na arbitragem, bem como a identificação de medidas tendentes ao aperfeiçoamento dessa intervenção. Sendo este um processo que para além do aspeto classificativo, tem implicações nas promoções e despromoções de categoria, devemos procurar a melhor compreensão sobre o instrumento utilizado, bem como do peso de cada fator para o processo de classificação final (Mascarenhas et al., 2005). A este propósito podemos referir que os observadores de árbitros assumem uma responsabilidade e uma exigência na sua atuação crucial para a classificação dos árbitros de futebol e consequentemente o desenvolvimento da sua carreira (Fruchart & Carton, 2012; Mathers & Brodie, 2011).

Tendo por base as exigências da atividade de arbitragem, mas também o contexto em que estas se desenvolvem, Mascarenhas et al., (2005) verificam que a performance do árbitro resulta da sua capacidade individual para intervir e gerir os fatores contextuais, que marcam o contexto de jogo, e o ambiente que o rodeia, e os fatores situacionais determinados pela dinâmica do jogo, e os aspetos disciplinares que o constituem. Não sendo possível estudar, em articulação, a avaliação entre estes 3 fatores e o modo como concorrem para a avaliação final do árbitro de futebol, é possível avaliar o modo como os fatores contextuais, que caraterizam os jogos e o ambiente que os rodeia, e os fatores situacionais, definidos pelo carater disciplinar da ação do árbitro na sua avaliação final.

Como referido anteriormente, os resultados obtidos revelam que, os fatores contextuais, apresentam valores percentuais mais elevados para o grupo dos árbitros melhor classificados, em comparação com os pior classificados, nomeadamente no que diz respeito ao número de jogos realizados por equipas da I Liga, aos jogos em que participaram os três primeiros classificados do campeonato e aos jogos que terminaram em empate. Já os árbitros que terminaram pior classificados apresentaram maior percentagem de jogos realizados por equipas da II Liga, jogos em que participaram outras equipas que não as três primeiras classificadas, e jogos desequilibrados.

De um modo geral considera-se que a atuação do árbitro, bem como a sua tomada de decisão pode ser afetada pelo contexto de jogo, nomeadamente pela avaliação do público e comunicação social, treinadores, dirigentes e jogadores (Collina, 2003). No entanto, de acordo com os resultados apresentados, os melhores árbitros parecem não sofrer tanto esta pressão do contexto do jogo, e possíveis consequências da avaliação do público e comunicação social sobre a sua atuação ou imagem (Catteeuw et al., 2009; Buraimo et al., 2012). Sabendo que os melhores tendem a arbitrar os jogos de contextos tendencialmente mais difíceis, é coerente que o processo de avaliação reflita essa mesma capacidade ao permitir que estes arbitrem mais jogos da I Liga, onde participam as equipas de maior impacto social (três primeiros classificados) e os jogos mais equilibrados. No entanto, face ao peso significativo que estas variáveis apresentam para a classificação final (cerca de 40%), consideramos que a classificação final pode ser enviesada logo à partida, pelas nomeações realizadas para estes jogos. Nomeadamente porque de acordo com os critérios de avaliação de desempenho do árbitro de futebol em vigor, um jogo com uma das equipas Top 3, é considerado automaticamente como um jogo de dificuldade mais elevado, aumentando de imediato a classificação final do árbitro, em comparação com um árbitro que tenha o mesmo desempenho, mas num jogo entre outros, ou entre equipas da II Liga.

De forma interessante, constatamos que dos quatro últimos árbitros na classificação final, e que desceram de categoria, apenas o último classificado, foi nomeado para mais jogos da I Liga (12) do que da II Liga (10). Do mesmo modo, o último classificado foi nomeado para dois jogos com a participação de equipas Top 3, enquanto os restantes três árbitros não foram nomeados para nenhum jogo Top 3. Estes resultados realçam que, apesar da tendência verificada ao nível das variáveis contextuais, existem casos em que o peso das variáveis relacionados com a ação mais técnica do árbitro (intervenção no jogo, gestão do jogo e aplicação das leis do jogo) poderão ter um peso significativo na sua classificação, bem como a articulação entre ambas é evidente. Por exemplo, um fraco desempenho num jogo terá reflexos diretos na classificação do árbitro no final da época desportiva, bem como indiretamente terá repercussões sobre o nível de jogos em futuras nomeações, o que se poderá aparentemente tornar um fator negativo, para a sua classificação final.

Estudos futuros são necessários para avaliar o peso de cada uma das variáveis das grelhas de avaliação, na classificação final do jogo, bem como da época desportiva de um árbitro de futebol. Dessa forma será possível identificar com maior precisão qual a influência destas variáveis contextuais na avaliação da prestação de cada árbitro de futebol.

Por sua vez, estudos anteriores têm revelado que a gestão dos aspetos disciplinares por árbitros peritos, em comparação com árbitros de menor experiência, nomeadamente a amostragem de cartões amarelos ou vermelhos (Fruchart & Carton, 2012) poderão ser influenciadas por "regras não escritas" (Plessner & Raab, 1999) permitindo uma gestão do jogo mais eficaz e de acordo com o contexto em que este ocorre (Askins, 2001; Grunska, 2001), em detrimento do estrito cumprimento das leis de jogo (Weaver, 2001). No entanto, os resultados obtidos, não nos revelam qualquer efeito ao nível do número de cartões amarelos e vermelhos exibidos por jogo. A este propósito, provavelmente, mais do que o número de cartões exibidos, a classificação final será mais influenciada pela avaliação, da intervenção disciplinar, do que pela quantidade de intervenções em si mesma. Por outro lado, ao invés dos árbitros de nível inferior, que tendem a mostrar mais cartões amarelos, aumentando o grau de dificuldade do jogo e consequentemente a classificação no mesmo em termos de dificuldade de gestão do jogo (Fruchart & Carton, 2012), os árbitros analisados constituíam o quadro de árbitros da 1ª categoria, sendo todos considerados como árbitros peritos, e a arbitrar as Ligas mais competitivas e mediáticas em Portugal. De acordo com o referido, estudos futuros deverão considerar mais do que o número de cartões amarelos e vermelhos mostrados, mas também a sua adequabilidade de acordo com a aplicação das leis do jogo, bem como o número de infrações técnicas e ocorrências disciplinares registadas no decorrer do jogo (Catteeuw et al., 2009).

Em suma, os resultados alcançados, evidenciaram por um lado, que o número de ocorrências relacionadas com fatores situacionais do jogo, possuem uma influência residual na classificação final do árbitro de futebol, enquanto que os fatores contextuais, têm uma influência significativa, sobre a classificação e avaliação de um árbitro no final da época desportiva, contribuindo para a subida (ou indicação à FIFA) de categoria do árbitro, a sua manutenção ou para a descida de categoria. No global, apenas considerando as variáveis contextuais e situacionais o modelo explica cerca de 40% da classificação dos árbitros de futebol no final da época desportiva. Em termos práticos, tal significa, que por si só o contexto do número de jogos que arbitra, poderá influenciar em 40% a sua classificação final. Face ao verificado, são necessários estudos futuros, no sentido de averiguar qual o peso que cada critério de classificação tem na classificação final do árbitro, de modo a averiguar se existe correspondência com o determinado em termos teóricos, para avaliação do desempenho do árbitro de futebol. Como limitação deste estudo apontamos o facto de considerar apenas uma época desportiva. Para uma melhor compreensão do peso, sobretudo das variáveis contextuais, na classificação final dos árbitros de futebol numa época desportiva,

estudos com amostras superiores devem ser considerados, de modo a uma maior generalização dos resultados.

Capítulo 6 – Discussão Geral

Ao árbitro de futebol é exigido um desempenho de excelência (Cuskelly & Hoye, 2013), estando o seu desempenho sob um forte escrutínio dos clubes, média e adeptos (Johansen & Haugen, 2013). De facto, o desempenho do árbitro é crucial para a dinâmica e resultado do jogo, pois um único erro (penalti, fora de jogo duvidoso) pode afetar decisivamente o resultado do mesmo (Halsey, 2007) ou comprometer a verdade desportiva. A tarefa do árbitro é de enorme exigência, obrigando a constantes julgamentos e tomadas de decisão sobre o desempenho de terceiros, de forma rápida e muitas vezes com limitada informação (Lex et al., 2015).

De modo semelhante, a avaliação do árbitro de futebol constitui-se de enorme exigência, pois, para além de obrigar ao julgamento sobre julgamentos e tomada de decisão sobre terceiros, necessita considerar essa avaliação à luz da dinâmica e enquadramento de cada uma das decisões e ações do árbitro para uma adequada gestão do jogo (Mathers & Brodie, 2011; Pina et al., 2019; Russell et al., 2018). No entanto, pouco ou nada se sabe sobre este processo de avaliação, nomeadamente, sobre a relação entre cada indicador de desempenho avaliado e as avaliações finais do desempenho do árbitro, ou mesmo sobre o modo como o contexto ou o recurso a diferentes meios e perspetivas de avaliação influencia a avaliação do desempenho do árbitro de futebol pelo OA (Halsey, 2007).

Face ao exposto, os objetivos da presente investigação foram: (i) compreender o nível de concordância entre os diferentes métodos de observação e avaliação do desempenho dos árbitros de futebol, realizada pelos OA; (ii) avaliar a relação entre as variáveis de desempenho, avaliadas pelos OA e a sua avaliação no final do jogo; (iii) avaliar a relação entre a classificação dos árbitros de futebol de elite no final de uma época desportiva e as variáveis contextuais e situacionais dos jogos arbitrados. De maneira a alcançar os objetivos inicialmente delineados, foram efetuados três estudos de investigação, cujos principais resultados reforçam (i) a existência de diferentes níveis de concordância, no julgamento e avaliação do desempenho do árbitro face aos métodos de observação utilizados (ao vivo, ao vivo + vídeo e apenas vídeo); (ii) a identificação de diferentes relações entre as variáveis de desempenho avaliadas pelos OA de futebol e a avaliação obtida no final do jogo de futebol; (iii) a identificação de uma influência significativa dos fatores contextuais sobre a classificação final dos árbitros de futebol no final da época desportiva, ao contrário dos fatores situacionais do jogo que apresentaram uma influência residual.

A avaliação do desempenho do árbitro de futebol (Pina et al., 2019) é normalmente efetuada por ex-árbitros experientes, tendo por base a observação ao vivo ou em vídeo, fundamentada em grelhas de observação, onde constam as seguintes categorias de avaliação: i) dificuldade e intervenção no jogo e ii) ações disciplinares e gestão do jogo. Cada uma destas categorias de avaliação é constituída por um conjunto de indicadores de desempenho que permitem avaliar as diferentes dimensões que caraterizam o sucesso na arbitragem. Apesar do processo de avaliação estar bem definido e sustentado em diretivas muito claras e objetivas para os avaliadores, e sendo utilizado há bastantes anos por uma larga maioria de Federações de Futebol, não é claro qual o melhor método de observação (ao vivo, em vídeo ou a conjugação de ambos), bem como qual a relação existente entre a pontuação dos diferentes critérios avaliados e a pontuação final da avaliação do árbitro.

Em relação ao recurso aos métodos de observação, as diretrizes da convenção de arbitragem da UEFA referiam, até há pouco tempo, que cada jogo deveria ser observado no estádio. No entanto, desde a época passada, foram introduzidas alterações a esta recomendação, passando a ser aconselhado a combinação de duas observações diferenciadas por diferentes observadores: uma ao vivo no estádio e outra através de vídeo. Por outro lado, por exemplo, a Federação Inglesa de Futebol realiza a observação do desempenho de árbitros de futebol apenas através de observação vídeo. De acordo com os resultados observados no estudo 1, podemos afirmar que o recurso a diferentes métodos de observação, nomeadamente ao vivo, e através de vídeo por observadores diferentes contribui para uma discrepância significativa em algumas variáveis observadas, podendo contribuir para um viés no processo de avaliação do desempenho do árbitro de futebol (Halsey, 2007). Fruto da restrição ou ampliação da informação tida em consideração para o processo de avaliação (Dohmen & Sauermann, 2016; Gilis et al., 2008; Halsey, 2007) permitida através dos métodos ao vivo, ou em vídeo, por diferentes observadores, o método ao vivo revelou valores superiores na avaliação de todas as variáveis com exceção das variáveis dificuldade e personalidade. O facto de estas variáveis obterem classificações menores ao vivo do que através de vídeo pode estar relacionado com a capacidade de observação e enquadramento na dinâmica do próprio jogo da avaliação do árbitro, pois sendo a dificuldade, bem como a personalidade, variáveis de análise subjetiva, necessitam de um maior enquadramento para avaliação, o que muitas vezes não é possível através de vídeo.

Face aos resultados obtidos no estudo 1, podemos concluir que o recurso a uma única perspetiva de análise tende a promover viés no processo de avaliação e, como tal, por exemplo, a opção da Federação Inglesa de Futebol não é recomendada. Do mesmo

modo, a opção recomendada pela UEFA revelou também não ser a mais apropriada para aumentar a precisão do processo de análise do desempenho do árbitro, pois é a combinação de métodos de observação que revela mais discrepância entre os valores de avaliação obtidos. Em suma, os resultados do estudo 1 apresentado no capítulo 3, sugerem a realização do processo de avaliação do árbitro de futebol através de um processo de avaliação combinado entre ao vivo e vídeo, a partir do qual o OA poderá corrigir através de vídeo e com maior detalhe o julgamento realizado ao vivo (Mathers & Brodie, 2011).

Como limitações do estudo, consideramos não ter sido possível identificar o método de análise mais rigoroso, para avaliar o desempenho dos árbitros, assim como observamos a necessidade de melhor compreender, se o período de oito semanas, que distou, entre a observação ao vivo e a observação em vídeo, é efetivamente o conveniente para a verificação da avaliação realizada ao vivo. Usualmente este é um processo que é realizado após o jogo, ou no dia seguinte ao jogo, sendo o envio da avaliação do OA remetido ao conselho de arbitragem até após 24h do jogo. Deste modo, será necessário avaliar no futuro qual o tempo que deve mediar o processo de avaliação em vídeo face ao momento de avaliação ao vivo e ao método a adotar (re-avaliação total do jogo ou apenas situações críticas e dúbias identificadas).

No estudo 2 apresentado no capítulo 4, tendo por base as grelhas de observação dos OA, e as variáveis de desempenho consideradas para avaliação do desempenho do árbitro de futebol, verificam-se diferenças no peso que cada uma dessas variáveis tem sobre a determinação da nota final do árbitro no final do jogo. Nomeadamente, as variáveis, personalidade, trabalho em equipa, aplicação negativa das leis do jogo, controle disciplinar e de gestão dos jogadores (negativo) revelam diferenças significativas entre os árbitros que obtiveram melhor classificação e os que revelam pior classificação no final do jogo. De forma interessante, no seguimento do estudo anterior, as variáveis, personalidade, aplicação negativa das leis do jogo, controle disciplinar e de gestão dos jogadores (negativo) revelam valores menores quando avaliadas ao vivo e a variável trabalho em equipa o inverso. Significa que a avaliação destas variáveis ao vivo tende a valorizar mais a sua aplicação de acordo com o contexto de jogo ao invés do que acontece na análise em vídeo, em que as avaliações tendem a ser mais negativas de acordo com o cumprimento estrito das leis de jogo (Mascarenhas et al., 2002).

Para além do referido, na definição de um modelo explicativo sobre a avaliação do desempenho do árbitro de futebol verifica-se que as avaliações do trabalho em equipa e

a aplicação negativa das leis de jogo explicam cerca de 60% da avaliação final do árbitro de futebol. De facto, estes resultados vão ao encontro de estudos anteriores, que realçam que a eficácia na tomada de decisão (Helsen & Bultynck, 2004) e o trabalho em equipa (Aoba et al., 2011) são fatores caraterizadores dos árbitros de elite. Face aos resultados obtidos, podemos afirmar que o desempenho do árbitro de futebol é claramente determinado pelo trabalho em equipa e pela aplicação negativa das leis de jogo, sendo 40% desta avaliação de desempenho resultado da conjugação das restantes oito variáveis avaliadas. Este resultado permite iniciar uma discussão sobre se o processo de avaliação e o modo como as variáveis de desempenho estão a ser consideradas e ponderadas para a obtenção da classificação final vão ao encontro do pretendido, ou se este processo poderá ser normalizado de modo à obtenção de um processo de avaliação mais preciso, tendo por base a ponderação de cada um dos critérios de avaliação considerados. Do mesmo modo, sugerimos que o processo de avaliação repense a utilização de algumas das variáveis que não apresentam diferenças entre os árbitros com pontuação final superior ou pontuação final inferior, tais como dificuldade do jogo, condição física, ações disciplinares, aplicação positiva de leis de jogo e controlo disciplinar e gestão dos jogadores (positiva). No entanto, consideramos os resultados aqui apresentados correspondem apenas ao início da compreensão deste processo e que, para o futuro, deverão ser considerados estudos que permitam uma análise mais detalhada sobre o impacto de algumas variáveis contextuais, como resultado do jogo, agressividade das equipas, bem como as variáveis situacionais, número de ações disciplinares, número de faltas, número de cartões, número de penaltis ou outras variáveis, de modo a identificar a influência destes fatores específicos no próprio desempenho do árbitro de futebol no decorrer de um jogo.

Como limitações deste trabalho, podemos considerar que apenas foi tida em consideração uma época desportiva da I Liga Portuguesa de Futebol. Deste modo, apesar de não ser possível generalizar o modelo apresentado, consideramos que ele reflete a tendência do modelo de avaliação do árbitro de futebol em Portugal. Do mesmo modo, a análise avaliou a relação entre variáveis de desempenho e a avaliação de desempenho final do árbitro em cada jogo, não se realizando nenhuma análise específica tendo em consideração cada árbitro ou a sua classificação no final da época desportiva.

No sentido de uma compreensão mais global do desempenho do árbitro de futebol no decorrer de uma época desportiva, e não sendo possível relacionar o desempenho jogo a jogo com a classificação final do árbitro, mas tendo por base as exigências do contexto em que estas se desenvolvem, o terceiro estudo apresentado no capítulo 5, procurou

uma melhor compreensão sobre a influência dos fatores contextuais e situacionais dos jogos arbitrados por cada árbitro na sua classificação final. De acordo com Mascarenhas et al., (2005), a performance do árbitro resulta da sua capacidade individual para intervir e gerir os fatores contextuais que marcam o contexto de jogo e o ambiente que o rodeia e como tal, estes devem ser considerados quando pretendemos avaliar o seu desempenho. Tendo por base a avaliação/classificação do árbitro de futebol no final de uma época desportiva, verifica-se uma influência significativa de algumas variáveis contextuais na classificação final do árbitro, mas não das variáveis situacionais consideradas, relacionadas com a ação disciplinar do árbitro no decorrer do jogo. Os árbitros melhor classificados no final da época desportiva foram os que arbitraram mais jogos de competição de nível superior (I Liga), realizaram mais jogos com equipas Top 3 e mais jogos com resultados equilibrados quando comparados com os árbitros com pior classificação. O modelo proposto tendo por base apenas a variação nas três variáveis contextuais referidas explica cerca de 40% da classificação dos árbitros de futebol no final de uma época desportiva.

Os resultados apresentados vão ao encontro de estudos anteriores, onde se revela que a atuação do árbitro, bem como a sua tomada de decisão podem ser afetadas pelo contexto de jogo (Collina, 2003), sendo que os árbitros de elite parecem não sofrer tanto esta pressão do contexto do jogo com evidentes melhorias na avaliação de desempenho obtida e na própria imagem face aos restantes intervenientes (Buraimo et al., 2012; Catteeuw et al., 2009). De facto, pode existir até uma relação com os resultados do segundo estudo, sendo que a capacidade de trabalho em equipa e a capacidade de tomar melhores decisões parecem ficar mais comprometidas em árbitros de menor nível, sobretudo quando constrangidos por fatores externos, como sejam a competitividade do jogo, o nível das equipas ou mesmo o equilíbrio entre elas (Collina, 2003). Os melhores árbitros tendem a arbitrar os jogos de contextos tendencialmente mais difíceis, e é coerente que o processo de avaliação reflita essa mesma capacidade, ao permitir que estes arbitrem mais jogos da I Liga, onde participam as equipas de maior impacto social (três primeiros classificados) e os jogos mais equilibrados (Fruchart & Carton, 2012). Apesar dessas evidências, tal como anteriormente, gostaríamos que estes resultados permitissem iniciar uma discussão sobre o peso que estes fatores contextuais parecem ter na classificação final dos árbitros de futebol. De facto, se estes fatores explicam cerca de 40% da classificação final, será que o processo de nomeação para cada jogo não poderá desde logo enviesar a classificação final dos árbitros de futebol? Especificamente, as variáveis contextuais apresentam uma influência significativa sobre a classificação e avaliação de um árbitro de futebol no

final da época desportiva, contribuindo para a subida (ou indicação à FIFA) de categoria do árbitro, a sua manutenção ou para a descida de categoria. Face ao exposto, são necessários mais estudos que permitam uma maior compreensão sobre este processo, nomeadamente, na compreensão e relação entre fatores contextuais e o desempenho efetivo de cada árbitro no jogo e ao longo da época desportiva. Uma maior compreensão nesta relação permitirá potenciar a identificação dos árbitros mais aptos em função das caraterísticas dos jogos em avaliação. No entanto, consideramos que o processo de avaliação ao longo da época desportiva deverá minimizar o impacto das variáveis contextuais na avaliação final do árbitro de futebol.

Em suma, a concretização deste trabalho permite afirmar que:

i) Existem diferentes níveis de concordância no julgamento e avaliação do desempenho do árbitro entre os métodos de observação ao vivo, ao vivo + vídeo e apenas vídeo. O uso de diferentes contextos de observação pode causar um impacto crítico na pontuação final do desempenho do árbitro no jogo. Os resultados sugerem ainda que a combinação da análise ao vivo e em vídeo permite aos observadores dos árbitros estarem mais seguros dos acontecimentos, que transmitem informações relevantes, para uma melhor avaliação do desempenho do árbitro.

ii) As variáveis de desempenho avaliadas pelos observadores dos árbitros têm pesos diferentes para a avaliação do desempenho do árbitro no jogo. O modelo é responsável por 60,4% da explicação da pontuação da avaliação do árbitro do jogo, com base nas variáveis do trabalho em equipa e da aplicação negativa das leis do jogo.

iii) Os fatores situacionais do jogo analisados possuem uma influência residual na classificação do árbitro de futebol no final de uma época desportiva, enquanto que os fatores contextuais apresentaram uma influência significativa sobre a classificação e avaliação de um árbitro no final da época desportiva, contribuindo para a subida (ou indicação à FIFA) de categoria do árbitro, a sua manutenção, ou para a descida de categoria. No global, apenas considerando as variáveis contextuais e situacionais o modelo explica cerca de 40% da classificação final dos árbitros de futebol. Em termos práticos, tal significa que por si só o contexto do número de jogos que arbitra poderá influenciar em 40% a sua classificação no final de uma época desportiva.

O trabalho desenvolvido na presente tese oferece evidências para o redesenho dos procedimentos da observação oficial e das políticas dos órgãos dirigentes do futebol, responsáveis pela avaliação no jogo e classificação final dos árbitros de futebol.

Capítulo 7 – Estudos futuros

Tendo por base os resultados alcançados, existe ainda muito para saber sobre qual o método de avaliação, bem como sobre a definição do instrumento para a avaliação do árbitro do jogo e classificação do árbitro de futebol no final da época desportiva.
Existe a necessidade da definição de um método e de um instrumento de avaliação, que corresponda, às necessidades do sector da arbitragem em particular e ao futebol em geral. Assim sendo indicamos, portanto, possíveis futuras investigações:

- Definir qual a metodologia a utilizar para a obtenção do cálculo da pontuação final, i.e., (i) cada uma das variáveis em análise contribui de forma independente para obter a pontuação final, ou (ii) ou existirá uma dependência de cada variável para a obtenção da pontuação final;
- Desenvolver estudos com o objetivo de avaliar e quantificar o peso e influência das diferentes variáveis (principalmente as situacionais), no processo da avaliação do árbitro, do jogo, e na sua classificação no final da época desportiva;
- Estudar os aspetos subjetivos da tomada da decisão, cuja avaliação, devido à sua subjetividade, pode ser avaliada pelo observador, como correta ou incorreta (exemplo: bola na mão ou mão na bola, qual a decisão acertada);
- Desenvolver a investigação, no sentido de encontrar um método de avaliação/classificação do árbitro de futebol, que responda de forma precisa para a análise de um dos diferentes contextos da avaliação;
- Definir se a avaliação de um árbitro deve ser apenas efetuada pelo observador, independentemente do método aplicado, ou deverá existir mais entidades que contribuam para a avaliação e classificação final;
- Identificar se a avaliação será mais influenciada pela avaliação da intervenção na área disciplinar, ou pela quantidade das intervenções disciplinares.

Capítulo 8 – Referências bibliográficas

Al Hazmi, A. (2016). *The movement and positioning of soccer officials in relation to their decision making*. Ph.D. Thesis, Queensland University of Technology, Autralia.

Aoba, Y., Yoshimura, M., Miyamori, T., & Suzuki, S. (2011). Assessment of soccer referee performance during games. *Football Science*, *8*, 8–15.

Araújo, D., Davids, K., Cordovil, R., Ribeiro, J., & Fernandes, O. (2009). How does knowledge constraint sport performance? An ecological perspective. In D. Araujo, H. Ripoll, & M. Raab (Eds.). *Perspectives on Cognition and Action in Sport,* (pp. 119–132). New York: Nova Science Publishers.

Askins, R. (2001). Common myths about officiating. *Referee, 26*(10), 44–47.

Ato, M., López, J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, *29*(3), 1038–1059. https://doi.org.10.6018/analesps.29.178511

Bar-Eli, M., Plessner, H., & Raab, M. (2011). *Judgement, decision making and success in sport. John Wiley & Sons,* Nova Jersey, EUA. https://doi.org/10.1002/9781119977032

Berry, J., Abernethy, B., & Cote, J. (2008). The contribution of structured activity and deliberate play to the development of expert perceptual and decision-making skill. *Journal of Sport & Exercise Psychology*, *30*(6), 685–708. https://doi.org/10.1123/jsep.30.6.685

Brand, R., Schmidt, G., & Schneeloch, Y. (2006). Sequential effects in elite basketball referees' foul decisions: An experimental study on the concept of game management. *Journal of Sport and Exercise Psychology, 28,* 93–99. https://doi.org:10.1123/jsep.28.1.93

Buraimo, B., Forrest, D., & Simmons, R. (2010). The 12th man?: Refereeing bias in english and german soccer. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 173*(2), 431–449. https://doi.org/10.1111/j.1467-985X.2009.00604.x

Buraimo, B., Simmons, R., & Maciaszczyk, M. (2012). Favoritism and referee bias in european soccer: Evidence from the spanish league and the UEFA champions league. *Contemporary Economic Policy, 30*(3), 329–343. https://doi.org/10.1111/j.1465-7287.2011.00295.x

Carvalho, V., Esteves, P., Nunes, C., Araújo, D., Helsen, W., & Travassos, B. (2020). Observe and make a call: Football referee's assessment is context sensitive. *International Journal of Performance Analysis in Sport, 20(6), 982–993.* https://doi.org/10.1080/24748668.2020.1820194

Catteeuw, P., Helsen, W., Gilis, B., & Wagemans, J. (2009). Decision-making skills, role specificity and deliberate practice in association football refereeing. *Journal of Sports Sciences*, *27*(11), 1125–1136. https://doi.org/10.1080/02640410903079179

Collina, P. (2003). *The rules of the game*. London: Macmillan

Cuskelly, G., & Hoye, R. (2013). Sports officials' intention to continue. *Sport Management Review*, *16*(4), 451–464. https://doi.org/10.1016/j.smr.2013.01.003

Dohmen, T., & Sauermann, J. (2016). Referee bias. *Journal of Economic Surveys*, *30*(4), 679–695. https://doi.org/10.1111/joes.12106

Dosseville, F., Laborde, S., & Bernier, M. (2014). Athletes' expectations with regard to officiating competence. *European Journal of Sport Science*, *14*(1), 448–455. https://doi.org/10.1080/17461391.2012.713006

Elsworthy, N., Burke, D., Scott, B., Stevens, C., & Dascombe, B. (2014). Physical and decision-making demands of australian football umpires during competitive matches. *Journal of Strength and Conditioning Research*, *28*(12), 3502–3507. https://doi.org:10.1519/JSC.0000000000000567

Elsworthy, N., Burke, D., & Dascombe, B. (2014). Factors relating to the decision-making performance of Australian football officials. *International Journal of Performance Analysis in Sport, 14*(2), 401–410. https://doi.org:10.1080/24748668.2014.11868730

Erikstad, M., & Johansen, B. (2020). Referee bias in professional football: Favoritism toward successful teams in potential penalty situations. *Frontiers in Sports and Active Living*, *2*(19). https://doi.org/10.3389/fspor.2020.00019

Folkesson, P., Nyberg, C., Archer, T., & Norlander, T. (2002). Soccer referees' experience of threat and agression: effects of age, experience, and life orientation on outcome of coping strategy. *Aggressive Behavior, 28*, 317–327. https://doi.org/10.1002/ab.90028

Fleiss, J., Cohen, J., & Everitt, B. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin, 72*(5), 323. https://doi.org/10.1037/h0028106

Fruchart, E., & Carton, A. (2012). How do amateur soccer referees destabilize a match? *Psicólogica: International Journal of Methodology and Experimental Psychology, 33*(3), 435–449.

Garicano, L., Palacios-Huerta, I., & Prendergast, C. (2005). Favoritism under social pressure. *The Review of Economics and Statistics, 87*(2), 208–216. https://doi.org:10.1162/0034653053970267

Ghasemi, A., Momeni, M., Rezaee, M., & Gholami, A. (2009). The difference in visual skills between expert versus novice soccer referees. *Journal of Human Kinetics*, *22*, 15–20. https://doi.org/10.2478/v10078-009-0018-1

Gilis, B., Helsen, W., Catteeuw, P., & Wagemans, J. (2008). Offside decisions by expert assistant referes in association football: Perception and recall of spatial positions in complex dynamic events. *Journal of Experimental Psychology*: *Applied, 14*(1), 21–35. https://doi.org/10.1037/1076-898X.14.1.21

Grunska, J. (2001). Yeah, You're Good…But You Can Be Better. *Referee 26*(6), 36–40.

Guillén, F., & Jiménez, H. (2001). Características deseables en el arbitraje y el juicio deportivo. *Revista De Psicologia Del Deporte*, *10*(1), 23–34.

Guillén F., & Deborah, F. (2011). A conceptual model of referee efficacy. *Frontiers in Psychologic, 2*(25), 1-5. https://doi.org/10.3389/fpsyg.2011.00025

Halsey, L. (2007). Travesties of Justice: The noise to signal ratio in association football. *Soccer & Society*, *8*(1), 68–74. https://doi.org/10.1080/14660970600989475

Helsen, W., Starkes, J., & Hodges, N. (1998). Team sports and the theory of deliberate practice. *Journal of Sport and Exercise Psychology*, *20*, 12–34. https://doi.org/10.1123/jsep.20.1.12

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

Helsen, W., Gilis, B., & Weston, M. (2006). Errors in judging "offside" in association football: Test of the optical error versus the perceptual flash-lag hypothesis. *Journal of Sports Sciences, 24*(5), 521–528. https://doi.org/10.1080/02640410500298065

Johansen, B., & Haugen, T. (2013). Anxiety level and decision-making among Norwegian top-class soccer referees. *International Journal of Sport and Exercise Psychology, 11*(2), 215–226. https://doi.org/10.1080/1612197X.2013.773665

Jones, M., Paull, G., & Erskine, J. (2002). The impact of team's aggressive reputation on the decisions of association football referees. *Journal of Sport Sciences, 20*(12), 991–1000. https://doi.org/10.1080/026404102321011751

Kittela, A., Larkina, P., Elsworthy, N., &, Spittle, M. (2019). Video-based testing in sporting officials: A systematic review. *Psychology of Sport & Exercise, 43*, 261–270. https://doi.org/10.1016/j.psychsport.2019.03.013

Lago-Peñas, C., & Gómez-López, M. (2016). The influence of referee bias on extra time in elite soccer matches. *Perceptual and Motor Skills, 122*(2), 666–677. https://doi.org/10.1177/0031512516633342

Lawrence, I., & Lin, K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*(1), 255-268. https://doi.org/10.2307/2532051

Lex, H., Pizzera, A., Kurtes, M., & Schack, T. (2015). Influence of players' vocalisations on soccer referees' decisions. *European Journal of Sport Science, 15*(5), 424–428. https://doi.org/10.1080/17461391.2014.962620

MacMahon, C., Starkes, J., & Deakin, J. (2007). Referee decision making in a video-based infraction detection task: application and training considerations. *International Journal of Sports Science & Coaching*, *2*(3), 257–265. https://doi.org/10.1260/174795407782233164

Mallo, J., Frutos, P., Juárez, D., & Navarro, E. (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), 1437–1445. https://doi.org/10.1080/02640414.2012.711485

Marôco, J. (2014). *Análise de Equações Estruturais*. (2nd ed.). Pêro Pinheiro, Report Number. Retrieved from http://www.reportnumber.pt/aee/

Mascarenhas, D., Collins, D., & Mortimer, P. (2002). The art of reason versus the exactness of science in elite refereeing: Comments on Plessner and Betsch (2001). *Journal of Sport and Exercise Psychology*, *24*, 328–333. https://doi.org/10.1123/jsep.24.3.328

Mascarenhas, D., Collins, D., & Mortimer, P. (2005). Elite refereeing performance: Developing a model for sport science support. *The Sport Psychologist*, *19*, 364–379. https://doi.org/10.1123/tsp.19.4.364

Mather, G. (2008). Perceptual uncertainty and line-call challenges in professional tennis. *Proceedings of the Royal Society B. Biological Sciences*, *275,* 1645–1651. https://doi.org/10.1098/rspb.2008.0211

Mathers, J., & Brodie, K. (2011). Elite refereeing in professional soccer: A case study of mental skills support. *Journal of Sport Psychology in Action*, *2*(3), 171–182. https://doi.org:10.1080/21520704.2011.609018

Morris, G., & Connor, D. (2016). Key attributes of expert NRL referees. *Journal of Sports Sciences*, *35*(9), 852–857. https://doi.org/10.1080/02640414.2016.1194524

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

Pietraszewskia, P., Maszczyka, A., Rocznioka, R., Gołaśb, A., & Stanula, A. (2014). Differentiation of Perceptual Processes in Elite and Assistant Soccer Referees. In *Procedia Social and Behavioral Sciences,* 3rd *International congress on physical education, sport and kinetotherapy (icpesk), Natl Univ Phys Educ & Sport, Bucharest, Romania*, jun 14-15, 2013*, 117*, (469–474). https://doi.org/10.1016/j.sbspro.2014.02.247

Pina, J., Passos A., Carvalho, H., & Maynard, M. (2019). To be or not to be an excellent football referee: Different experts' viewpoints. *Journal of Sports Sciences, 37*(6), 692–700. https://doi.org/10.1080/02640414.2018.1522940

Plessner, H., & Raab, M. (1999). Kampf- und schiedsdchterurteile als produkte sozialer Informations-verarbeitung [Judgments by officials in sports as products of social information processing]. *Psychologie & Sport, 6*, 130–145.

Plessner H., & Betsch, T. (2001). Sequential effects in important referee decisions: The case of penalties in soccer. *Journal of Sport & Exercise Psychology*, *23*, 254–259. https://doi.org/10.1123/jsep.23.3.254

Put, K., Wagemans, J., Pizzera, A., Williams, A., Spitz, J., Savelsbergh, G., & Helsen, W. (2016). Faster, slower or real time? Perceptual-cognitive skills training with variable video speeds. Psychology of Sport and Exercise, *25*, 27–35. http://doi.org/10.1016/j.psychsport.2016.03.007

Russell, S., Renshaw, I., & Davids, K. (2018). How interacting constraints shape emergent decision-making of national-level football referees. *Qualitative Research in Sport, Exercise and Health*, 1–16. https://doi.org/10.1080/2159676X.2018.1493525

Sabatini, D. (2002). The search for excellence. *Referee,* 24–28.

Schnyder, U., & Hossner, E. (2016). Psychological issues in football officiating: An interview study with top-level referees. *Current Issues in Sport Science*, *1*, 1–14. https://doi.org/10.15203/CISS_2016.004

Schweizer, G., Plessner, H., Kahlert, D., & Brand, R. (2011). A video-based training method for improving soccer referees' intuitive decision-making skills. Journal of Applied Sport Psychology, *23*(4), 429–442. https://doi.org/10.1080/10413200.2011.555346

Sülün, Ö. (2013). *Comparison of anger and furiousness levels of soccer referees with their empathetic tendencies*. Master's thesis. Karamanoğlu Mehmet Bey University, Institute of Social Sciences, Karaman.

Sutter, M., & Kocher, M. (2004). Favoritism of agents–the case of referees' home bias. Journal of Economic Psychology, 25(4), 461–469. https://doi.org/10.1016/S0167-4870(03)00013-8

Tenreiro-Gavela, F., Montero-Seone, A., & Saavedra-García, M. (2016). La autopercepción del esfuerzo y recuperación en el arbitraje profesional español: Un estudio de caso. *Cuadernos De Psicología Del Deporte*, *16*(3), 137–144.

Unkelbach, C., & Memmert, D. (2008). Game management, context effects, and calibration: The case of yellow cards in soccer. *Journal of Sport & Exercise psychology, 30*(1), 95–109. https://doi.org/10.1123/jsep.30.1.95

Urra, B., Nuñez, C., Oses, J., & Sarmiento, G. (2018). Variables psicológicas influyentes en el desempeño de árbitros FIFA: Un estudio cualitativo. *Revista de Psicología Aplicada al Deporte y al Ejercicio Físico*, *3*(2), 1–11. https://doi.org/10.5093/rpadef2018a13

Weaver, P. (2001). No hiding place for Gallagher. The guardian unlimited. Monday November 12th. Electronic article retrieved from http://football.guardian.co.uk/Match_Report/0,591738,00.html

Weston, M., Castagna, C., Impellizzeri, F., Bizzini, M., Williams, A., & Gregson, W. (2012). Science and medicine applied to soccer refereeing: An update. *Sports Medicine, 42*(7), 615–631. https://doi.org/10.2165/11632360-000000000-00000

Williams, M., & Ericsson, A. (2005). Perceptual-cognitive expertise in sport: Some considerations when applying the expert performance approach. *Human Movement Science*, *24*(3), 283–307. https://doi.org/10.1016/j.humov.2005.06.002

Wilson, R. (Producer & Director) (2000, July 25). *Man in black*. Scotland: British Broadcasting Corporation.

Wolfson, S., & Neave, N. (2007). Coping under pressure: Cognitive strategies for maintaining confidence among soccer referees. *Journal of Sport Behavior*, *30*, 232–247.

Anexos

Grelha de Avaliação (Época 2016-17)

	FEDERAÇÃO PORTUGUESA DE FUTEBOL
	CONSELHO DE ARBITRAGEM
	RELATÓRIO DE OBSERVAÇÃO(época 2016-2017)

Equipa A	Resultado Final __/___	**Equipa B**

Jornada:	Nº do Jogo:	Data:	Local:

1- Identificação de árbitros e observador:

Grau de Dificuldade

FUNÇÃO	NOME	ASSOCIAÇÃO	NOTA	1P	2P
ÁRBITRO					
AA 1					
AA 2					
QUARTO A					
OBSERVADOR					

Escala de Avaliação

9,00	a	10,00	Excelente desempenho	7,80			Insatisfatório
8,50	a	8,90	Muito Bom	7,50	a	7,70	Muito insatisfatório
8,30	a	8,40	Bom	7,00	a	7,40	Fraco
8,00	a	8,20	Satisfatório(+)	-7,00			Muito Fraco
7,90			Satisfatório(-)				

Grau de dificuldade

1= Normal	2= Difícil	3= Muito Difícil

2- Ação Disciplinar

MIN	EQUIPA	Nº	CARTÃO	CLASSIFICAÇÃO MOTIVO

3-Descrição do Jogo

Tempo adicional: 1T ___ 2T__ 1PP___ 2PP__

A- Árbitro

A1- APLICAÇÃO E INTERPRETAÇÃO DAS LEIS DO JOGO, CONTROLO E ABORDAGEM DO MESMO

Comentários:

Descrição das situações

Tempo	(+ ou -)	

A2- CONTROLO DISCIPLINAR, GESTÃO DE JOGADORES E REPRESENTANTES DAS EQUIPAS

Comentários:

Descrição das situações

Tempo	(+ ou -)	

A3- PERSONALIDADE

Comentários:

Descrição das situações

Tempo	(+ ou -)	

A4- CONDIÇÃO FÍSICA E MOVIMENTAÇÃO/COLOCAÇÃO

(+)	Expectável	(-)	
❑	❑	❑	MOVIMENTAÇÃO E ACOMPANHAMENTO DO JOGO
❑	❑	❑	POSICIONAMENTO EM SITUAÇÕES DE BOLA PARADA
❑	❑	❑	DIAGONAL ENQUADRAMENTO COM OS ÁRBITROS ASSISTENTES
❑	❑	❑	CONDIÇÃO FÍSICA

Descrição das situações

Tempo	(+ ou -)	

A5- COLABORAÇÃO ENTRE A EQUIPA DE ARBITRAGEM

(+)	Expectável	(-)	
❑	❑	❑	EFICIENTE COOPERAÇÃO E DISTRIBUIÇÃO DE RESPONSABILIDADES
❑	❑	❑	EFICIENTE COOPERAÇÃO EM SITUAÇÕES DIFÍCEIS
❑	❑	❑	EFICIENTE COMUNICAÇÃO/SINALÉTICA E CONSULTAS

Descrição das situações

Tempo	(+ ou -)	

A6- COMENTÁRIO FINAL DE ATRIBUIÇÃO DE NOTA E CONSELHOS

Comentários:

A7- PONTOS IMPORTANTES DO DESEMPENHO E ASPETOS A MELHORAR	
Pontos Positivos	
1	
2	
3	
Pontos a Melhorar	
1	
2	
3	

A8- Informação adicional

B- **Árbitro Assistente 1**

B1- DESEMPENHO DO ÁRBITRO ASSISTENTE

Comentários:

Descrição das situações

Tempo	(+ ou -)	

ASPECTOS RELEVANTES

(+)	Expectável	(-)	
❑	❑	❑	DECISÕES RELATIVAS A SITUAÇÕES DE FORA DE JOGO
❑	❑	❑	POSICIONAMENTO E MOVIMENTAÇÃO/COLOCAÇÃO
❑	❑	❑	COOPERAÇÃO COM O ÁRBITRO
❑	❑	❑	SINALÉTICA
❑	❑	❑	FALTAS NA ZONA DE AÇÃO

B2- Informação adicional

C- Árbitro Assistente 2

C1- DESEMPENHO DO ÁBITRO ASSISTENTE

Comentários:

Descrição das situações

Tempo	(+ ou -)	

ASPECTOS RELEVANTES

(+)	Expectável	(-)	
❑	❑	❑	DECISÕES RELATIVAS A SITUAÇÕES DE FORA DE JOGO
❑	❑	❑	POSICIONAMENTO E MOVIMENTAÇÃO/COLOCAÇÃO
❑	❑	❑	COOPERAÇÃO COM O ÁRBITRO
❑	❑	❑	SINALÉTICA
❑	❑	❑	FALTAS NA ZONA DE AÇÃO

C2- Informação adicional

D- **QUARTO ÁRBITRO**

D1- QUARTO ÁRBITRO

Comentários:

ASPECTOS RELEVANTES

(+)	Expectável	(-)	
❑	❑	❑	COOPERAÇÃO COM O ÁRBITRO
❑	❑	❑	PROCEDIMENTOS: NAS SUBSTITUIÇÕES/ ENTRADAS DAS EQUIPAS MÉDICAS E OUTROS
❑	❑	❑	ATITUDE PERANTE OS COMPORTAMENTOS DOS ELEMENTOS DOS BANCOS
❑	❑	❑	CONTROLO DE EQUIPAMENTOS E INDICAÇÃO DE TEMPO ADICIONAL
❑	❑	❑	ATENÇÃO E CONCENTRAÇÃO DEMONSTRADA

D2- Informação adicional

Data:

Assinatura_____