Decision-making Assessment and Development in Australian Football Umpires: Evaluation of 360° VR

Aden Brett Kittel

Bachelor of Exercise and Sports Science (University of Newcastle)

Bachelor of Applied Science (Honours) (Human Movement) (Victoria University)

College of Sport and Exercise Science

Institute for Health and Sport

This thesis is submitted in fulfilment of the requirements for admission to the degree of Doctor of Philosophy

Of

Victoria University

September 2020

Abstract

Many skills underpin the performance of sporting officials, however decision-making is regarded as the most critical. There are finite on-field opportunities to develop decision-making of sporting officials in training and competition, as a consequence, video-based approaches are typically used to assess and develop decision-making skill. Existing methods such as the use of match broadcast video may not be an ecologically valid method to present decision-making scenarios. With technological advancements, using virtual reality may improve the ecological validity of video-based approaches to improve decision-making. *Study 1* systematically reviewed existing research utilising video-based testing to assess decision-making in officials, which often differentiates between skill levels to demonstrate construct validity. *Study 1* identified several limitations including common use of match broadcast video, limited reporting of reliability, and studies often solely reporting number of decisions rather than performance accuracy. Comparison between video-based and in-game decision-making performance was rarely conducted. This study provided the foundation to further examine the efficacy of video-based tests in sporting officials. *Study 2* developed two valid and reliable video-based tests, based on the recommendations of *Study 1*. As match broadcast video is the most common video-based testing method for officials, it was compared with 360° VR to assess decision-making accuracy. Both 360° VR and match broadcast video-based tests demonstrated construct validity and high reliability (r = 0.89). Stronger ecological validity was evident in 360° VR than match broadcast, as participants rated 360° VR to be more representative of in-game decision-making processes. *Study 3* aimed to determine the relationship between decision-making accuracy in both video-based tests (360° VR and match broadcast) and in-game of elite Australian football umpires, given that this limitation of the research was identified in *Study* 1. *Study 3* used validated video-based tests from *Study 2*. There were no significant relationships observed for decision-making accuracy between in-game and video-based testing. *Studies 2* and *3* provide findings on testing, however it is unclear whether 360° VR or match broadcast is more effective for developing decision-making. *Study 4* assessed the effectiveness of a video-based training program using 360° VR or match broadcast to develop decision-making in amateur Australian football umpires using a randomised control study design. Decision-making was assessed

using the valid and reliable tests of *Study 2* before, immediately following, and one month following training (retention test). The 360° VR group exhibited significantly higher decision-making accuracy ($p < 0.05$) than the control group at retention testing, with no between-group differences observed for the match broadcast group. Participants rated 360° VR as more relevant and enjoyable than match broadcast. In summary, this thesis aimed to develop and evaluate the effectiveness of 360° VR as a video-based testing and training tool in Australian football umpires. Although 360° VR and match broadcast appear to have strong construct validity and reliability, currently, there is limited transfer to in-game performance. Further, based on these results, it is not definitive whether 360° VR is a more effective training tool than match broadcast. The findings of this thesis indicate 360° VR may be more ecologically valid than match broadcast and warrants further investigation.

Student Declaration

"I, Aden Brett Kittel declare that the PhD thesis by Publication entitled 'Decision-making Assessment and Development in Australian Football Umpires: Evaluation of 360° VR' is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work".

Signature:  **Date**: 2/9/2020

Aden Brett Kittel

List of Publications

Peer Reviewed Journal Publications

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

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. (2019c). Using 360° virtual reality as a decision-making assessment tool in sport. *Journal of Science and Medicine in Sport*, 22(9), 1049-1053. https://doi.org/10.1016/j.jsams.2019.03.012

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. (2020). Transfer of 360° virtual reality and match broadcast video-based tests to on-field decision-making. *Science and Medicine in Football*. https://doi.org/10.1080/24733938.2020.1802506

Kittel, A., Larkin, P., Elsworthy, N., Lindsay, R., & Spittle, M. (2020). Effectiveness of 360° virtual reality and match broadcast video to improve decision-making skill. *Science and Medicine in Football*. https://doi.org/10.1080/24733938.2020.1754449

Conference Presentations

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. The Validity & Reliability of 360° VR. Oral Presentation at the Australasian Skill Acquisition Network Conference, Sydney, Australia 2018.

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. 360º VR as a Decision-Making Tool for AFL Umpire Decision-Making. Oral Presentation at the World Congress on Science and Football, Melbourne, Australia 2019. *Best Student Investigator Award.*

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. 360º VR as a Decision-Making Tool for AFL Umpire Decision-Making. Oral Presentation at the FEPSAC, Munster, Germany 2019.

VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

PART A:

Details of Included Papers: Thesis by Publication

Please list details of each Paper included in the thesis submission. Copies of published Papers and submitted and/or final draft Paper manuscripts should also be included in the thesis submission

Item/ Chapter No.	Paper Title	Publication Status (e.g., published, accepted for publication, to be revised and resubmitted, currently under review, unsubmitted but proposed to be submitted)	Publication Title and Details (e.g., date published, impact factor etc.)
Ch.3	Video-based testing in sporting officials: A systematic review.	Published	**Kittel, A.**, Larkin, P., Elsworthy, N., & Spittle, M. (2019d). Video-based testing in sporting officials: A systematic review. *Psychology of Sport and Exercise*, 43, 261-270. https://doi.org/10.1016/j.psychsport.2019.03.013 Accepted 29/3/2019. Impact factor: 2.827
Ch.4	Using 360° virtual reality as a decision-making assessment tool in sport.	Published	**Kittel, A.**, Larkin, P., Elsworthy, N., & Spittle, M. (2019c). Using 360° virtual reality as a decision-making assessment tool in sport. *Journal of Science and Medicine in Sport*, 22(9), 1049-1053. https://doi.org/10.1016/j.jsams.2019.03.012 Accepted: 31/3/2019. Impact factor: 3.607
Ch.5	Transfer of 360° virtual reality and match broadcast video-based tests to on-field decision-making.	Accepted for publication	**Kittel, A.**, Larkin, P., Elsworthy, N., & Spittle, M. (2020). Transfer of 360° virtual reality and match broadcast video-based tests to on-field decision-making. *Science and Medicine in Football*. https://doi.org/10.1080/24733938.2020.1802506 Accepted: 6/7/2020. Impact factor: N/A
Ch.6	Effectiveness of 360° virtual reality and match broadcast video to improve decision-making skill.	Accepted for publication	**Kittel, A.**, Larkin, P., Elsworthy, N., Lindsay, R., & Spittle, M. (2020). Effectiveness of 360° virtual reality and match broadcast video to improve decision-making skill. *Science and Medicine in Football*. https://doi.org/10.1080/24733938.2020.1754449 Accepted: 6/4/2020. Impact factor: N/A

Declaration by: Aden Kittel **Signature:** [redacted] **Date: 2/9/2020**

Acknowledgements

Pursuing a PhD would not be possible without the assistance, guidance and support of many people. It is impossible to put into words the efforts of so many individuals in helping me over this journey.

Firstly, thank you to my three supervisors; Dr. Paul Larkin, Dr. Nathan Elsworthy, and Associate Professor Michael Spittle. You have all assisted in your own individual ways, and your steady support and encouragement has been invaluable.

Paul, thank you for time, energy and encouragement over this journey. Our regular chats about footy have always been a welcome distraction. Our journey to Germany together in 2019 is one of my favourite highlights and I hope we get another opportunity to have a beer at the Neckarmüller. I am proud to call you a mate, not just a supervisor.

Nathan, thank you for starting me on this academic journey. You set me up when I first arrived in Melbourne, and this thesis would never have existed without your ideas and our discussions from the start. I highly doubt I would be in Melbourne without your encouragement all the way back in 2015.

Michael, thank you for initially taking me on as an unknown Honours student back in 2016. Your expertise and mentorship has been vital for me in not only this thesis, but also my teaching pursuits. I will always appreciate your advice from the beginning to take ownership of my studies.

Dr. Sharna Spittle, thank you for the opportunities both at VU and overseas. These opportunities have often provided a welcome break from the thesis, but also helped me identify my love for teaching.

Thank you also to Dr. Christopher Stevens for being a key reason for me going into research. Not only were the opportunities and experience with you vital for me starting a PhD, but I am happier for how they led to the friendship we have today.

Research would not be possible without participants and key stakeholders. I would like to make special mention of Rowan Sawers. Your support for me since I moved to Melbourne has been

exceptional. Your enthusiasm for umpiring is infectious, and it has been a privilege to work with you. I look forward to catching up over a coffee soon (my shout). Thank you also to Steve Keating for your energy and your group's involvement. It was great to re-connect with you since our time in Sydney.

Thank you to the AFL Umpiring Department for their involvement and support of my project. In particular, thank you to Robert Jackson (Jacko) for your support of my research (particularly to get me started), but always to me as a person. Thank you also to Damian Farrow for your expertise and guidance throughout. Jacko and Damo – both of you have been great mentors, but your support of me when I needed it will always be appreciated. Thank you to Jacob Adams for your assistance.

One of the best parts of doing a PhD is the friends you make along the way. In particular, thank you to Pete Le Noury, Victoria Brackley, Riki Lindsay and Peter Browne for always making VU a place that I looked forward to coming to. I hope we get to celebrate all of us being called "Doctor" in the near future.

Mum, Dad and Talia. It hasn't been easy at times moving to a different state, for any of us. But you have all always supported me in whatever I do. Thank you for always motivating to have a strong work ethic and love for what you do. Words can't describe how much you all mean to me, and I would not be where I am today without the three of you. I hope we can celebrate with a drink together in person soon.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

When participating in sport, there are a range of factors or skills that contribute to successful performance. Perceptual-cognitive skills encompass the ability to identify crucial information in the sporting environment, incorporate this information with existing knowledge to select, and execute an appropriate response (Marteniuk, 1976). Perceptual-cognitive skills such as decision-making, anticipation, pattern recall and situational probabilities are a central component of sporting performance, with the sensitivity to differentiate between lesser and higher skilled performers (Berry, Abernethy, & Côté, 2008; Williams & Ericsson, 2005). Decision-making is one of the primary perceptual-cognitive skills in sport, defined as perceiving information, correctly interpreting, and selecting an appropriate response from the options perceived (Baker, Côté, & Abernethy, 2003). Scientific research investigating decision-making in sport has typically focused on decision-making skill in athlete populations (Baker, Cote, & Abernethy, 2003; Williams & Ericsson, 2005), with sports officials under-investigated in comparison.

There has been an increase in empirical research of officials in recent years (Hancock, Bennett, Roaten, Chapman, & Stanley, 2020). Interactor officials, such as soccer referees and Australian football umpires, defined as having high perceptual and physical demands (MacMahon et al., 2014), have frequently been examined in the literature. This developing research area is due to the important role officials play in sport, where each decision has the potential to influence the outcome of a match (Larkin, Berry, Dawson, & Lay, 2011). Given the official's important match-play role of enforcing the laws of the game, decision-making is considered the most important skill for successful performance (Helsen & Bultynck, 2004; Kittel, Larkin, Elsworthy, & Spittle, 2019b; Morris & O'Connor, 2016). Despite the importance of decision-making, more scientific study (Hancock et al., 2020) and training time (MacMahon, Helsen, Starkes, & Weston, 2007a) has been devoted to other components of performance such as physical fitness. Although there is a heavy focus on physical factors in the research and training environment (Raab, MacMahon, Avugos, & Bar-Eli, 2020), existing literature suggests that cognitive

attributes such as decision-making are much more important than physical fitness for successful performance (Kittel et al., 2019b; Morris & O'Connor, 2016).

The decision-making processes of Australian football umpires are very unique. For example, there are laws in Australian football where umpires must make decisions using certain processes/heuristics where they are several options for a potential free infringement of the laws (Larkin, Mesagno, Berry, & Spittle, 2018a). One of the most complex laws for an umpire to interpret is that for a tackling situation, with decisions such as 'holding the ball' are unique to the sport of Australian football (Larkin et al., 2018). Australian football umpires are also expected to apply the laws in a black and white manner, irrespective of the context surrounding the decision (Kittel et al., 2019b). This is different to other sporting officials such as soccer referees who help shape the game by being proactive in warning the players to encourage fairness, and sometimes take context into their decision-making process (Russell, Renshaw & Davids, 2019).

In terms of both assessing and training the critical skill of decision-making, it has been acknowledged in-game performance is the best method to measure (Bonney, Berry, Ball, & Larkin, 2019) and develop this skill (MacMahon et al., 2007a). There are several limitations of using only on-field performance measures to assess decision-making skill level, despite in-game performance being the most valid or "gold standard" measure (Bonney et al., 2019). For example, there are large fluctuations from game to game in terms of teams officiated, number and type of decisions, and crowd size to name a few (Corrigan, Dwyer, Harvey, & Gastin, 2019); which makes analysis and comparison of in-game performance difficult. In order to reliably assess decision-making performance, researchers and practitioners need to explore alternative reproducible measures to assess decision-making skill to use in conjunction with (not in replacement of) in-game processes.

A common off-field method to assess and develop decision-making skill is using video-based approaches, which present sport-specific decision-making scenarios in a video format with the goal of simulating on-field decision-making. Video-based approaches typically isolate the decision-making component (Larkin, Mesagno, Spittle, & Berry, 2015) to provide a more accurate understanding of an individual's skill level, negating the interference of these additional performance attributes. For the

purpose of assessing decision-making skill, video-based testing overcomes the limitation (i.e., game to game variability) of inferring decision-making skill from match performance. This is achieved by presenting consistent decision-making scenarios in video format, enabling a reliable tool to distinguish between-participant skill level and individual changes over time such as following a training intervention. No research currently exists which summarises the existing literature base of video-based testing in sports officials. As outlined by van Biemen, Koedijker, Renden, & Mann (2018), video-based approaches for testing and training decision-making typically decouple perception and action. When making a decision while using video-based methods, participants typically verbalise their decision or press a button. This is a key limitation when using video-based methods for athletes who typically perform perceptual-motor skills, such as making a decision then passing the ball. Officials perform perceptual-cognitive tasks where they do not pass or intercept a ball. As this is more representative of actions in games, sports officials may then be an ideal group to examine video-based methods which do not couple perception and action.

For video-based tests to produce meaningful and useful outcomes, it is important to establish the reliability and validity of these measures (Larkin et al., 2015). Video-based testing has been investigated as a means to distinguish between expert and novice decision-makers. For example, video-based testing demonstrate construct validity by differentiating between expert and novice Australian football players (Lorains, Ball, & MacMahon, 2013b) and umpires (Larkin, Mesagno, Berry, & Spittle, 2014a). This concept of construct validity (i.e., distinguishing between different performance levels) is common in video-based testing research examining the differences between expert and novice decision-makers (Gadotti, Vieira, & Magee, 2006). There are, however, several limitations present in the existing video-based testing literature. For example, few studies report the reliability of the video-based test (Larkin et al., 2014a; Spitz, Put, Wagemans, Williams, & Helsen, 2017), which is an important consideration to maximise the robustness and reproducibility of a testing tool. In addition, it has been recommended that researchers should examine the transfer of performance in off-field decision-making tasks to on-field performance which would provide further evidence of their validity (Paradis, Larkin, & O'Connor, 2016). Researchers commonly overlook these data due to the difficulty of obtaining such

sensitive information, and the lack of control researchers have over the game environment leading to possible reliability issues. The comparison of off-field to on-field decision-making performance is integral to determine the validity of using both traditional and novel video-based approaches.

In addition to being used as an assessment tool, video-based approaches are more commonly implemented as a means of developing decision-making skill. Given the popularity of this research area, several reviews have been published outlining the current state of video-based training research and future directions to be addressed (Broadbent, Causer, Williams, & Ford, 2015; Larkin et al., 2015; Renshaw et al., 2018). Previous studies use a range of training methods such as varied instructional approaches (Farrow & Abernethy, 2002), development of decision-making under fatigue (Kittel, Elsworthy, & Spittle, 2019a), and manipulation of video speeds (Lorains, Ball, & MacMahon, 2013a). Given the difficulty of obtaining first-person video footage, a common video presentation method is termed match broadcast footage (footage generally used for televised games) which is filmed from a fixed, elevated position in the grandstand (Larkin, Mesagno, Berry, Spittle, & Harvey, 2018b; Lorains et al., 2013a). This is the most common video mode in video-based training studies in officials (Kittel et al., 2019a; Larkin et al., 2018b; Schweizer, Plessner, Kahlert, & Brand, 2011). Mann, Farrow, Shuttleworth & Hopwood (2009) assessed perceptual-cognitive differences between first-person and an elevated perspective similar to match broadcast, with players making better decisions in the elevated perspective. Officials, on the other hand, make different types of decisions to players. For example, players make decisions on who to pass to, using information such as open space and the relative position of different players (Mann et al., 2009) to dictate their decisions. Officials, however, often make decisions based on the physical contact between two players (Larkin et al., 2018) and therefore do not use the same cues such as spatial information of relevant players. There are limitations of using this footage, however, as it presents a third-person perspective filmed from a sideline position rather than first-person perspective, thus limiting the perceptual information perceived in actual competition (Craig, 2013). This, therefore limits the representativeness (i.e., similarity to real competition) and ecological validity (i.e., similarity of perceptual cues) of the task (Araujo, Davids, & Passos, 2007). A key component of a representative task is maintaining sources of information in the simulation that are

relevant to performance, in order for expert advantages to be present, leading to a more valid tool (Dicks, Button, & Davids, 2010; Ericsson, Krampe, & Tesch-Römer, 1993).

To overcome the limitations associated with match broadcast footage as both a training and assessment tool, researchers have attempted to increase the representativeness of video-based tools by implementing first-person viewpoints (Petit & Ripoll, 2008). A key concept of representativeness is the visual correspondence of the task, or how similar the perceptual information presented is to that of a game (Pinder, Davids, Renshaw, & Araújo, 2011). Although first-person footage increases this first aspect (visual correspondence) of task representativeness, behavioural correspondence (mimicking the movement pattern of match play) is also integral (Pinder et al., 2011). Virtual reality, defined as simulations of a real or imaginary environment for a participant to perceive and move/interact within the environment, has been suggested as a way to overcome the above limitations (Craig, 2013). Virtual reality increases the behavioural correspondence, or perception-action coupling of the task, whereby athletes complete a sporting action in response to the decision-making stimulus such as moving to intercept an opposition pass in rugby union (Brault, Bideau, Kulpa, & Craig, 2012). Virtual reality can be presented in a number of ways, such as a head mounted display, 3D projection, or a room with screens surrounding the participant to present an interactive experience (Craig, 2013). Head mounted displays are an effective means to increase behavioural correspondence of a task compared to footage presented on a flat screen, as the video perspective automatically updates with changes in head movements (Craig, 2013). Initial studies in virtual reality used animated environments and characters to present perceptual-cognitive scenarios (Brault et al., 2012; Vignais, Kulpa, Brault, Presse, & Bideau, 2015). These virtual approaches have been developed to assist with perception-action coupling, which is typically missing from existing screen-based approaches (Craig, 2013). Despite these positive advancements in the literature, representativeness may be increased by presenting real-world, rather than animated footage.

360° virtual reality (360° VR; also known as immersive video) has emerged as a novel technology which retains some of the strengths of animated virtual reality (first-person perspective, head movements), but using real-world 360° video presented on a head mounted display. As this

presents a first-person perspective, this may be a more ecologically valid tool. For off-field decision-making tasks, it is important these are representative of performance constraints, and present similar perceptual cues as the perceptual cues present in competition (i.e., ecological validity) (Araujo et al., 2007). Two studies have recently examined the effectiveness of this technology in developing decision-making in basketball players (Pagé, Bernier, & Trempe, 2019; Panchuk, Klusemann, & Hadlow, 2018), with results indicating that 360° VR can be effective in training decision-making skill in basketball players. Theoretically, 360° VR has stronger ecological validity than existing screen-based approaches, as the perceptual cues from the first-person perspective are more similar to those perceived in match play (Araujo et al., 2007). In contrast to virtual reality (i.e., virtual environments), 360° VR is a suitable training tool in populations who perform more perceptual-cognitive than perceptual-motor skills (i.e., does not require naturalistic interaction with the environment; e.g., pass or intercept) such as sports officials (Fadde & Zaichkowsky, 2018). Although studies have suggested 360° VR is more ecologically valid than existing methods (Bird, 2020; Fadde & Zaichkowsky, 2018), there are no previous studies which have quantified the game-likeness of different video-based approaches such as 360° VR and match broadcast.

There are several gaps in the research to be addressed as this novel technology emerges both in the literature and practical settings. Despite the importance to reliably assess decision-making skill in off-field settings for sports officials, no research currently exists summarising the existing video-based testing literature in sports officials. Secondly, the validity and reliability of existing match broadcast video and novel 360° VR technology should be established to identify the reproducibility and robustness of each for decision-making assessment. Although 360° VR is theoretically more ecologically valid than match broadcast, no research exists which quantifies the ecological validity of each video mode. A reliable and valid tool will enable the assessment of performance change over time, such as following a training intervention within sports officials. Thirdly, more research is required to determine the relationship of off-field (both existing match broadcast and 360° VR) approaches to on-field decision-making skill. This is pertinent in sports officials where few studies have examined this comparison. Theoretically, skilled perceptual-cognitive performance in representative tasks typically transfers to in-

game, therefore it is necessary to assess this. If a test can be developed with a strong relationship to on-field performance, this would assist with reliable talent identification and/or selection based on decision-making skill. Finally, although some studies have suggested 360° VR may be an effective decision-making training tool, this technology should be contrasted to existing video-based training methods to examine the real effect of 360° VR compared to traditional methods. Representative learning design suggests practice activities which are closer to performance demands will lead to stronger learning adaptations. More research is required to assess the efficacy of more representative approaches such as 360° VR. Given the importance of decision-making skill to officiating performance, the use of 360° VR may establish a novel method to assess and develop decision making performance within sports officiating cohorts. No such research has examined the efficacy of these methods within Australian football officials.

1.1. Structure and aims of the thesis

Broadly, this thesis aimed to develop and evaluate the effectiveness of 360° VR as a video-based testing and training tool in Australian football umpires' decision-making skill. While there are a range of perceptual-cognitive skills, given its importance to performance in sporting officials, such as Australian football umpires, decision-making will be the skill investigated in this thesis. Although this thesis focuses on Australian football umpires, it is anticipated that the findings will be transferrable not only to officials of different sports, but also to athletes. This thesis begins with the literature review (Chapter 2), followed by the four studies of this thesis: Study 1 (Chapter 3), Study 2 (Chapter 4), Study 3 (Chapter 5), and Study 4 (Chapter 6). Finally, Chapter 7 provides a summary of the research and future recommendations. This thesis is submitted as a thesis by publication. The following studies have been published or accepted for publication:

- Chapter 3 (Kittel, Larkin, Elsworthy, & Spittle, 2019d) was published in *Psychology of Sport and Exercise*;
- Chapter 4 (Kittel, Larkin, Elsworthy, & Spittle, 2019c) was published in *Journal of Science and Medicine in Sport*;
- Chapter 5 (Kittel, Larkin, Elsworthy, & Spittle, 2020) has been accepted for publication in *Science and Medicine in Football*;
- Chapter 6 (Kittel, Larkin, Elsworthy, Lindsay, & Spittle, 2020) has been accepted for publication in *Science and Medicine in Football*.

As decision-making is the most important skill for sports officials, studies have become more prevalent in recent years attempting to understand this skill and the factors that influence it (Corrigan et al., 2019; Paradis et al., 2016; Spitz, Moors, Wagemans, & Helsen, 2018). The investigation of on-field decision-making is the most valid method to explore this skill, though there are limitations of on-field decision-making assessment. For example, there are only a finite amount of games to officiate, each of which have high physical loads and there is a high degree of variability from game to game. Therefore, off-field video-based approaches are commonly used to examine decision-making skill.

Study 1 (Chapter 3) will present a systematic review of the existing knowledge base of video-based testing in sporting officials, with the specific aims to:

i. Summarise video-based decision-making assessment literature in the domain of interactor officials.
ii. Analyse the various methods utilised to simulate match-like decision-making.

Several studies have identified the efficacy of video-based testing in assessing decision-making skill of athletes and officials in a controlled, off-field manner, as summarised in *Study 1*. With the advent of technology, 360° VR is becoming widely accessible as a tool that may provide a more game-like presentation of decision-making scenarios than previous methods. *Study 2* (Chapter 4) contributes to existing knowledge by validating this novel technology along with previously used match broadcast footage. The specific aims of *Study 2* are to:

i. Examine both the reliability and construct validity of two video-based methods to assess Australian football umpires' decision-making skill. These two video-based methods are 360° VR (i.e., first-person viewpoint) and traditional match broadcast footage (i.e., third-person viewpoint).
ii. Evaluate the level of ecological validity of each method through game-likeness ratings.

Although numerous studies have utilised video-based tests to measure decision-making skill as evident in *Study 1*, very few have provided a comparison of performance in these tests to decision-making accuracy in a match (i.e., competitive environment). This is an important consideration as a test that accurately reflects in-game performance is a tool which has sensitivity to identify future talent, discriminate between participants in a controlled manner, and provide benchmarks for new and existing athletes or officials for selection purposes. Therefore, the aim of *Study 3* is to:

i. Determine whether two reliable and valid video modes (360° VR and match broadcast vision) can distinguish between elite participants, based on game decision-making performance.

Video-based training programs are a common method for athletes and officials to develop decision-making skill to complement in-game practice. Studies (see Larkin et al. (2015) for a review) have demonstrated that video-based decision-making interventions can hasten the development of decision-making skill, particularly in amateur participants. Most video-based training studies utilise match broadcast video (third-person viewpoint). The aims of *Study 4* are:

i. Determine the effectiveness of both 360° VR and traditional match broadcast video-based training methods in the development of decision-making skill of amateur Australian football umpires.
ii. Examine the representativeness of each video mode through assessment of psychological fidelity and perception of deliberate practice.

Chapter 7 will present a short summary of each study outlined above, along with conclusions and recommendations arising from this research. Given the results described herein, future research directions will be outlined.

It is anticipated this research will provide a significant contribution to existing research investigating tools to assess and develop decision-making skills in interactor sports officials, with these findings transferrable to athlete populations also. As there recently has been an increase in research into the efficacy of 360° VR, these findings will contribute to existing knowledge on the suitability of this technology as an assessment and training tool. The efficacy and effectiveness of this technology will be directly compared to previously used match broadcast video.

Chapter 2: Literature Review

2.1. Officiating in sport

2.1.1. General roles and importance to sport

In any competitive team sport, there are typically two teams opposing one another. Matches are played in accordance with a set of rules and regulations to ensure a fair and safe environment for players. The officials, commonly known as the umpires or referees, are ultimately responsible for implementing the laws of the game. As such, officials play an integral role in sporting competition and the importance of this role is highlighted by the potential impact a correct or incorrect free kick/decision can have on the game's outcome (Larkin et al., 2011).

2.1.2. Types of officials

There are various classifications of sporting officials according to the specific role they have and type of scenarios they assess. MacMahon et al. (2014) developed a classification system of three categories of officials; including *reactors*, *monitors*, and *interactors* (Figure 2.1). Officials are categorised based on the interaction with athletes, and the number or athletes or cues to monitor. *Reactors* have low interaction with their environment and cues to monitor, whereby they typically respond to a single event (i.e., tennis line judge). While still having limited interaction with the environment, *monitors* have a higher number of cues and/or athletes to monitor such as judging a performance (i.e., gymnastics judge). This review, however, will focus on *interactors* who are defined as having high perceptual demands and interaction with their environment (MacMahon et al., 2014). Team sport officials such as Australian football umpires, soccer referees, and rugby league referees are all examples of interactors.

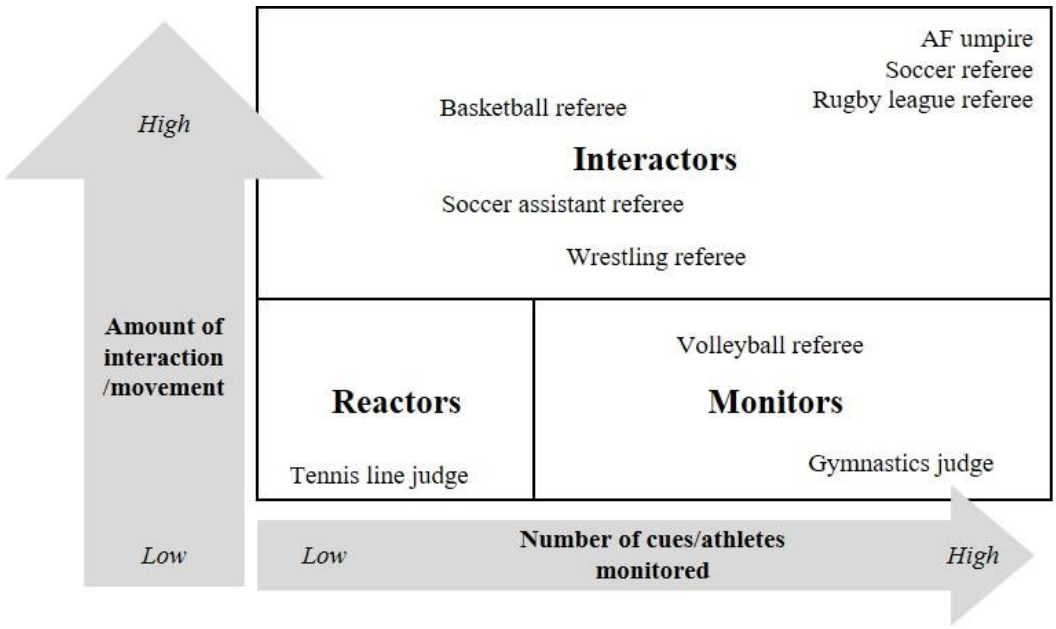

Figure 2.1: Categories of officials from MacMahon et al. (2014).

Note: This continuum is based upon the number of athletes/cues monitored by the official, in conjunction with the interaction and movement completed.

2.1.2.1. Interactor officials

Interactor officials are defined as having extensive physical and perceptual loads (MacMahon et al., 2014). For example, the physical loads of officials are extensive, covering distances such as approximately 10,000m in soccer (Weston et al., 2012), 7,000 - 9,000m in rugby league (Emmonds et al., 2015), 8,000m in rugby union (Blair, Elsworthy, Rehrer, Button, & Gill, 2018), and Australian football umpires covering 10,500 – 12,000m per game (Elsworthy, Burke, Scott, Stevens, & Dascombe, 2014; Elsworthy & Dascombe, 2011). In conjunction with the physical demands, sporting officials are required to continually perform perceptual decision-making tasks, based upon the actions of the players. Previous studies have reported soccer officials make on average 137 observable decisions per game, with a significantly higher number of non-observable decisions (Helsen & Bultynck, 2004). Specific to Australian football umpires, Elsworthy et al. (2014) reported that these officials award on average 44 free kicks (i.e., penalties) per game (not including other decisions such as marks). In addition to free kicks, Australian football umpires are suggested to view upwards of 2,000 decision moments per match following an analysis of communication strategies (Neville, Salmon, & Read, 2016). These moments

include marks, non-observable and borderline decisions not considered a penalty. It is evident that the different types of interactor officials have significant physical and perceptual loads, as indicated by MacMahon et al. (2014). Given the substantial physical and perceptual loads, fitness and perceptual-cognitive skills appear to be important for successful officiating for interactor officials (Kittel et al., 2019b; Morris & O'Connor, 2016).

2.1.3. Characteristics of elite officiating

Interactor officials such as those from soccer, rugby, and Australian football possess a range of attributes that contribute to performance success. Mascarenhas, Collins, and Mortimer (2005b) developed a framework of components that contribute to successful rugby refereeing performance, with transferability of this model to other interactor officials. This research identified four cornerstones of successful officiating including *physical fitness, positioning & mechanics; knowledge & application of the law; contextual judgement; and personality & game management skills;* all overarched by the *psychological characteristics of excellence*. Similarly in soccer referees, there are a range of factors that contribute to performance excellence. These include mental toughness attributes, support networks and services, effective game-management qualities, multi-faceted pre-match preparation, performance-level enhancement, opportunities to thrive, personal characteristics, and superior physical components (Slack, Maynard, Butt, & Olusoga, 2013). These studies provide a general snapshot of the range of attributes required for successful officiating performance in interactor officials.

To understand key factors indicating successful performance in sporting officials, it is valuable to rank these to identify the most important attributes. Morris and O'Connor (2016) developed a hierarchy of attributes required for rugby league officiating excellence. This knowledge provides direction for coaches to allocate more training time and researchers a framework of which skills need greater investigation to improve performance. This study identified decision-making, defined as a cognitive attribute to be the most important attribute for successful performance, with other cognitive attributes such as reading the game and communication ranked within the top three (Morris &

O'Connor, 2016). These findings demonstrated a high degree of importance placed on cognitive attributes, whereas fitness was the only attribute categorised as physical that officials ranked within the top ten attributes. This suggests some cornerstones of officiating performance, such as knowledge and application of the law (Mascarenhas et al., 2005b), are more instrumental for successful officiating than other cornerstones such as physical fitness.

A study specifically in Australian football has further identified the attributes required for Australian football umpiring performance as rated by elite umpires (Kittel et al., 2019b). Similar to rugby league referees (Morris & O'Connor, 2016), elite umpires rated cognitive attributes to be of utmost importance to Australian football officiating, with knowing the laws/interpretations and decision-making ranked as one and two, respectively (Kittel et al., 2019b). Knowledge of the laws is more declarative knowledge, whereas decision-making is the application of this as procedural knowledge. Applying this knowledge through decision-making may then be considered the most important observable skill for Australian football umpires. These two attributes form the key cornerstone of knowledge and application of the law developed by Mascarenhas et al. (2005b). The high ranking of knowledge of the laws could be credited to the complexity of officiating Australian football; where, for example, there are at least ten different decisions that could be made for a tackling situation (Larkin et al., 2018a). This knowledge is required for officials to make a decision accurately, highlighting the inextricable link between the top two attributes for Australian football umpires.

By understanding the relative importance of each attribute to successful performance, this provides guidance for structuring training programs. For example, as officials have commonly cited decision-making (i.e., knowledge and application of the law) as the fundamental performance attribute (Helsen & Bultynck, 2004; Kittel et al., 2019b; Morris & O'Connor, 2016), it should receive the greatest focus in training. The development of fitness, however, comprises a significant portion of officials' training time (Catteeuw, Helsen, Gilis, & Wagemans, 2009). Fitness was rated as the 14th most important attribute for Australian football umpiring performance (Kittel et al., 2019b), compared to 9th most important for rugby league referees (Morris & O'Connor, 2016). Despite this, several studies have solely investigated physical loads in Australian football umpires (Elsworthy & Dascombe, 2011;

Jessiman, Harvey, Corrigan, & Gastin, 2019), similar to rugby league (Brightmore et al., 2016; Emmonds et al., 2015) and soccer referees (see Weston et al. (2012) for a review). Additionally, studies have investigated the relationship between physical fatigue and decision-making accuracy in on-field (Elsworthy et al., 2014) and off-field (Paradis et al., 2016) domains, with results suggesting a negligible relationship between fatigue and decision-making. Therefore, researchers can focus on and isolate decision-making aspects of umpiring performance, rather than aspects that are not considered as integral for successful performance. In addition, given the importance of this skill, officials present a unique opportunity to study novel decision-making assessment and training procedures that can be used in a wider range of environments such as with athletes themselves.

2.1.4. Defining decision-making in sports officials

Perceptual-cognitive skills are defined as the ability to identify and acquire environmental information for integration with the existing knowledge such that appropriate responses can be selected and executed (Marteniuk, 1976). Decision-making is a key perceptual-cognitive skill in sport, as skilled decision-making precedes an appropriate and context specific action (O'Connor, Larkin, & Williams, 2017). Decision-making has been defined as the ability to plan, select, and execute an action based on the individual's knowledge in addition to information available in the current situation (Williams & Ford, 2013). Therefore, an ability to use appropriate game-play information to guide skilled movement is a fundamental component of performance (Abernethy & Russell, 1987; Williams, Davids, & Williams, 1999). From an athletic performance perspective, decision-making can be defined as the process of choosing the most appropriate movement response (e.g., passing to a teammate) from a range of possible options (Abernethy, 1996); or the selection of functional actions (e.g., taking a shot on goal) from a range of possible actions to achieve a specific goal (Hastie, 2001). Similar to athletes, decision-making remains a central component in sport officiating performance where perceptual and judgement processes are central to effective decision-making (Morris & O'Connor, 2016). Sports officials' decision-making can be broadly defined as applying and interpreting the laws of the game in a particular

sport, by identifying whether a player has infringed (Morris & O'Connor, 2016). This is to ensure that players play the game in a fair and safe manner.

2.1.5. On-field decision-making of sports officials

As highlighted by MacMahon et al. (2014), interactor officials have a large amount of perceptual cues and players to monitor as part of their role. In Australian football for example, there are 36 players on the field at any time, with umpires requiring expert perceptual decision-making skills to officiate effectively (Larkin et al., 2011). From a sporting officials' perspective, an official makes a decision every time they see an athlete infringe upon the laws of the game. When making a decision such as a free kick/penalty, officials make a signal using their hand or arms, and/or blow a whistle. While this action accompanies the decision, accuracy is not a critical component of this action such as a player executing a pass. In addition to awarding free kicks/penalties, there are many more situations where officials make deliberate 'no penalty' or 'play on' decisions (Neville et al., 2016). The game (competition) data/match statistics do not necessarily record or quantify these decisions not to award a penalty and allow the game to continue, but they form a significant portion of the decision-making load in sporting officials. It may, therefore, be difficult to reliably assess decision-making performance in-game due to 'no penalty' decisions not included in the assessment. When evaluating decision-making, decisions are typically classified as 'hits', 'misses' or 'false alarms (van Biemen et al., 2018). In Australian football umpire studies, 'hits' are defined as 'correct', 'misses' as 'missed', and 'false alarms' as 'unwarranted' (Elsworthy et al., 2014, Corrigan et al., 2019).

According to MacMahon et al. (2014), there are similarities in the decision-making cues and player monitoring requirements of interactor officials across sports. Common interactor officials (and the player total for each team to monitor) reported throughout the literature include soccer referees (11), rugby league referees (13), rugby union referees (15), and Australian football umpires (18). For each of the sports listed above, except for rugby union, previous studies have quantified the in-game decisions and the accuracy of officials. Table 2.1 indicates there is a considerable amount of decisions made per

game, reiterating the impact of the performance of officials within the sporting environment. Accuracy of decisions is usually determined based on correct decisions identified by subject matter experts, who are the officials' coaches (Corrigan et al., 2019; Elsworthy et al., 2014). The studies in Table 2.1 highlight the need for future decision-making development, with officials typically making higher than 20% incorrect decisions (Corrigan et al., 2019; Emmonds et al., 2015; Mascarenhas, Button, Hara, & Dicks, 2009).

Table 2.1: In-game decision-making demands of interactor officials.

Sport	***Study***	***Competition***	***Free kicks awarded (n)***	***Decision-making accuracy (%)***
Soccer				
	Helsen and Bultynck (2004)	UEFA (elite)	44	NR
	Mascarenhas et al. (2009)	NZ national league (sub-elite)	~21	64%
	Mallo, Frutos, Juárez, and Navarro (2012)	FIFA Cup (elite)	25 ± 3	86%
Australian football				
	Elsworthy et al. (2014)	AFL (elite)	44 ± 8	84 ± 6%
	Corrigan et al. (2019)	AFL (elite)	40 ± 7	78%
Rugby league				
	Emmonds et al. (2015)	NRL (elite)	15 ± 2	74 ± 5%

AFL: Australian Football League; UEFA: United European Football Associated; NZ: New Zealand; EPL: English Premier League; FIFA: Fédération Internationale de Football Association; NRL: National Rugby League.

There are numerous contextual factors that can influence decision-making of sporting officials, thus limiting the accuracy of decisions within competition. Firstly, given the high physical demands of interactor officials (MacMahon et al., 2014), researchers have commonly examined the influence of fatigue on in-game decision-making to mixed results. In Australian football umpires, decision-making accuracy does not waver over the course of a game, suggesting that fatigue does not influence this skill (Elsworthy et al., 2014). Alternatively, interactor officials such as soccer (Mallo et al., 2012) and rugby league referees (Emmonds et al., 2015) exhibit a decrease in accuracy in the last 10 to 15 minutes of match play. There are other external factors such as home advantage (Goumas, 2014), crowd size (Downward & Jones, 2007), margin between teams and position on the ground (Corrigan et al., 2019), which may impact decision-making processes of interactor officials. It has also been reported that more

experienced officials tend to be better decision-makers in competition (Corrigan et al., 2019). It is evident from the factors listed above that in-game decision-making can vary depending on the context in which decisions occur.

On-field decision-making assessment is the optimal method to assess decision-making skill in officials (MacMahon et al., 2007a), similar to athletes (Bonney et al., 2019). The assessment of on-field decision making, however, is somewhat difficult, given the subjective assessment from game to game. Naturally, not all umpires officiate the exact same game, which can limit the reliability of assessing multiple officials' performances. In addition, coaches grade decisions using match broadcast footage, which is not the same perspective of an official in-game, as it is captured from an elevated perspective in the grandstand (Craig, 2013). Given the variability of on-field decision-making, there is a need to develop measures to assess decision-making in officials in a controlled, reliable manner. Off-field methods overcome this limitation of on-field (within competition) assessment, as they provide stronger control and consistency than on-field assessment (Larkin et al., 2014a). In conjunction with on-field assessment, studies have suggested off-field decision-making assessment can be used for the purpose of talent identification and performance benchmarking (Larkin et al., 2014a) and repeated measures of skill development over time such as following a training intervention (Larkin et al., 2015).

2.2. Theoretical frameworks for off-field perceptual-cognitive tasks

2.2.1. Key frameworks

When designing off-field assessment and training tasks which simulate on-field scenarios, it is important these tasks represent the perceptual-cognitive (i.e., decision-making) processes that officials experience within competition. There have been numerous studies introducing theoretical frameworks and commentaries to facilitate the development of off-field tools for testing and training decision-making skill (Fadde & Zaichkowsky, 2018; Farrow, Reid, Buszard, & Kovalchik, 2018; Hadlow, Panchuk, Mann, Portus, & Abernethy, 2018). Two key theoretical frameworks underpinning this thesis were representativeness and ecological validity. Representativeness refers to the extent task constraints

in experimental designs (i.e., off-field tasks) represent the constraints experienced in the specific sporting/performance environment (Araújo, Davids, & Hristovski, 2006; Pinder et al., 2011). Ecological validity refers to the relationship between a cue (perceptual variable) to that of the performance environment (Araujo et al., 2007). Although these two terms are very similar, studies have reinforced the distinction between the two concepts (Araujo et al., 2007; Pinder et al., 2011). For example, representativeness refers to how factors (i.e., constraints) in an experimental setting represent those of competition, whereas ecological validity is how valid (i.e., similar to competition) the perceptual cues are in the experimental (off-field) setting (Araujo et al., 2007; Pinder et al., 2011). In general terms, both concepts refer to how closely an off-field task can represent on-field performance environments. It is important for off-field tasks to have high representativeness and/or ecological validity, as this will more likely elicit the expertise differences that would be present in actual competition (Farrow et al., 2018).

2.2.2. High fidelity simulations

For a task to be representative of the on-field competition setting, there are several key considerations for researchers. Firstly, the simulation needs to have high levels of fidelity, which refers to the extent a situation replicates reality and is a key element of transfer (Alessi, 1988; Farrow, 2013). Action fidelity refers to whether a performer's responses are similar between the experimental (i.e., off-field) and performance (i.e., on-field) setting (Pinder et al., 2011). A higher level of action fidelity can be achieved through perception-action coupling, where simulated scenarios require the participant to complete the action they would within a game (Craig, 2013). For example, this may require a pass for a rugby player (Correia, Araújo, Cummins, & Craig, 2012) or intercepting a shot on goal for a handball goalkeeper (Vignais et al., 2015). For optimal simulation of on-field performance, it is imperative that tasks have strong action fidelity by incorporating perception-action coupling (Bideau et al., 2010).

For sporting officials on the other hand, there is no accompanying motor action following the perceptual decision, as they verbalise their response (Neville, Salmon, & Read, 2018). Therefore,

psychological and physical fidelity would be more important for this cohort than action fidelity. Psychological fidelity refers to how life like the participant perceives the simulation to be, whereas physical fidelity refers to much the simulation looks like the real performance environment (Lorains et al., 2013a; Stoffregen, Bardy, Smart, & Pagulayan, 2003). Both forms of fidelity refer to how realistic a task is. Off-field tasks can increase fidelity by utilising first-person viewpoints, as opposed to third-person viewpoints, which provide an inaccurate representation of the perceptual information received in a game (Craig, 2013). This results in stronger visual correspondence, where the perceptual information received in a simulation is very similar to a game (Pinder, Headrick, & Oudejans, 2015). Previous research has found expert soccer players make faster and more accurate decisions in a first-person task (Petit & Ripoll, 2008), supporting the hypothesised expert differences of a representative task (Dicks, Davids, & Button, 2009).

2.2.3. Modified perceptual training framework for representative tasks

Recently, Hadlow et al. (2018) developed the Modified Perceptual Training Framework (MPTF) to assist researchers and practitioners in designing perceptual tasks which transfer to on-field performance. There are three key assumptions when using a tool for perceptual-cognitive skills, including; 1) it needs to be able to distinguish between skill levels, 2) improvements can be made through training, and 3) any performance changes can be assessed through a transfer to on-field performance (Hadlow et al., 2018). By meeting the three assumptions listed above, perceptual training tools such as video-based training will have the strongest transfer to the field. To create a stronger perceptual-cognitive testing or training tool, Hadlow et al. (2018) suggest these tools target high-order perceptual-cognitive skills such as sport-specific decision-making (Williams & Ford, 2008), rather than generic visual skills such as visual acuity (Erickson, 2007). In addition to the perceptual skill targeted, these tools must have strong visual correspondence (Pinder et al., 2015) and behavioural correspondence (Pinder et al., 2011), which are both integral components of representative tasks. When considering the different components of this framework, technologies such as flat-screen video have been used to present sport-specific stimuli with moderate visual correspondence. Tools such as virtual

reality are emerging, given the strong behavioural correspondence of this technology (Hadlow et al., 2018). These two technologies (i.e., existing video-based approaches and virtual reality) will be outlined throughout this literature review.

2.3. Video-based approaches for off-field decision-making testing and training

2.3.1. Off-field decision-making assessment

Off-field decision-making assessments have been used throughout the literature for both athletes (Lorains et al., 2013b) and officials (Larkin et al., 2014a) with the ability to distinguish between skill levels. This method is often titled "video-based testing" which uses video footage to present a sport-specific decision-making scenario, in order to apply knowledge of the sport's rules to provide an accurate decision (Larkin et al., 2014a; Mascarenhas et al., 2005b). Video-based testing in sports officials has examined decision-making from a variety of perspectives. For example, the effect of crowd noise (Balmer, Nevill, Lane, & Ward, 2007; Lex, Pizzera, Kurtes, & Schack, 2015; Nevill, Hemingway, Greaves, Dallaway, & Devonport, 2017), relationship between physical exertion and decision-making (Larkin et al., 2014b; Paradis et al., 2016), and expert-based differences (Ghasemi, Momeni, Jafarzadehpur, Rezaee, & Taheri, 2011; Larkin et al., 2011; Spitz, Put, Wagemans, Williams, & Helsen, 2016). There is a significant body of research utilising video-based technology to assess decision-making in sporting officials. Therefore, a systematic review of relevant studies was completed for Study 1 (Chapter 3) of this thesis. This systematic review examined the existing knowledge base of video-based testing in sporting officials, highlighted the limitations of existing methods, and presented recommendations for future research employing this method to assess decision-making.

2.3.2. Developing decision-making skills

Knowledge of the laws and decision-making are the two most important skills for Australian football umpiring performance (Kittel et al., 2019b). Therefore, knowledge and application of the law (Mascarenhas et al., 2005b) requires a strong training emphasis for Australian football umpires. In current training procedures, Australian football umpires spend a large amount of time in lecture-style meetings developing declarative law-based knowledge (Paradis et al., 2016), however, they require more training time for the application of this knowledge (i.e., decision-making). Officiating games has been recognised as the ideal method to train decision-making skill in soccer officials (MacMahon et al., 2007a). Further, the amount of in-game practice is positively correlated with skill in soccer officials (Catteeuw et al., 2009). However, there are only a finite amount of games to officiate with substantial physical loads (Elsworthy et al., 2014; Jessiman et al., 2019), therefore additional time and tools are required to further develop decision-making skill. As such, video-based training has emerged as a common tool for decision-making skill development, defined as presenting sport-specific scenarios in video format, requiring a specific response from an individual (Larkin et al., 2015). As these provide sport-specific decision-making scenarios, these methods can increase deliberate practice hours to hasten development of this skill (Catteeuw et al., 2009; MacMahon et al., 2007a). Deliberate practice was originally introduced by Ericsson et al. (1993) to describe activities completed in training that are done individually, effortful, guided by a coach and are not necessarily enjoyable. However, deliberate practice has been redefined by Starkes, Deakin, Allard, Hodges, & Hayes (1996) to be high in relevance, enjoyment, concentration and effort. This adjusted definition by Starkes et al. (1996) has been used in studies investigating sports officials (MacMahon et al., 2007a, Catteeuw et al., 2009).

2.3.3. Video-based training overview

Overall, video-based approaches are an effective technique to develop perceptual-cognitive skill (Larkin et al., 2015) in a range of athletic populations, such as invasive sports (Gabbett, Carius, & Mulvey, 2008; Gorman & Farrow, 2009), interceptive sports (Farrow & Abernethy, 2002; Hopwood,

Mann, Farrow, & Nielsen, 2011), and sports officials (Catteeuw, Gilis, Jaspers, Wagemans, & Helsen, 2010b; Larkin et al., 2018b; Schweizer et al., 2011). Studies of video-based approaches typically use a quasi-experimental design to determine the efficacy of the intervention program, with amateur/novice athletes the most commonly researched population group (Larkin et al., 2015). The training duration for effective video-based training varies greatly, where interventions may be as brief as a single session (Shafizadeh & Platt, 2012), or as extensive as 18 sessions over six weeks (Hopwood et al., 2011). An intervention totalling 4-8 weeks is considered sufficient for decision-making development in amateur participants (Larkin et al., 2015).

2.3.4. Video-based training studies in sports officials

Given the importance of decision-making skill for sporting officials (Helsen & Bultynck, 2004; Kittel et al., 2019b), video-based training has been examined as a method to improve this skill. These studies have been conducted with sporting officials in sports such as soccer (Schweizer et al., 2011), rugby union (Mascarenhas, Collins, Mortimer, & Morris, 2005c), and Australian football (Larkin et al., 2018b). As summarised in Table 2.2, there are a variety of video-based training approaches in sporting officials. There are, however, several similar features across studies, where most of the video-based training programs in sporting officials have utilised match broadcast footage (i.e., from an elevated perspective in the grandstand) as the preferred presentation mode (Kittel et al., 2019a; Larkin et al., 2018b; Schweizer et al., 2011; van Biemen et al., 2018). Mascarenhas et al. (2005c) used first-person video from competitive games in rugby union referees to emphasise the ecological validity of the training. The situations filmed were scrum scenarios, where an individual made their way onto the field to film from a similar perspective to where the official would be. These scenarios are more stationary in nature, and it would be logistically near impossible for someone to follow the play in more dynamic environments such as soccer and Australian football.

Table 2.2: Video-based training in interactor officials

Sport	Study	Video format	Instructional type	Training duration (no. sessions per week) - clips	Groups (participant no.)	Main effect	Conclusion
Australian football							
	Kittel et al. (2019a)	Broadcast	Explicit	8 weeks (1) - 96	Video before HIIT (6), Video during HIIT (7), Control (7)	No significant difference observed between groups.	Video-based training improvements were unclear following the intervention.
	Larkin et al. (2018b)	Broadcast	Implicit	12 weeks (1) - 1,040	Intervention group (21), control group (31)	Significant improvement in intervention group from pre to retention test ($p < 0.05$). Less experienced umpires in the intervention group significantly improved ($p = 0.01$).	Video-based training appears to be beneficial, especially for less experienced umpires.
Soccer (*central)*							
	van Biemen et al. (2018)	Broadcast	Explicit	1 week (1) - 70	Normal video (11), blurred video (11)	Normal group decreased close to significance ($p = 0.10$). Blur group increased close to significance ($p = 0.07$). Significant difference between groups ($p = 0.03$).	Blurred vision training appears to be more beneficial for DM improvement than normal training.
	Schweizer et al. (2011)	Broadcast	Explicit	7 weeks (1) - 144	With repetition (19), without repetition (19), control (10)	Both training groups significantly improved decision-making skill ($p < 0.05$).	Video-based training can be an effective method with immediate feedback to develop decision-making.

Table 2.2 (continued)

Sport	Study	Video format	Instructional type	Training duration (no. sessions per week) - clips	Groups (participant no.)	Main effect	Conclusion
Soccer (*assistant)*							
	Put et al. (2016)	1st person simulated – adult performers	Explicit	3 weeks (1) - 60	Increasing speed over each session 75% - 100% - 125% (33), decreasing speed 125% - 100% - 75% (33), arbitrary speed 100% - 75% - 125% (30)	Only decreasing speed group improved following intervention. All groups had significantly less flag errors.	If manipulating video speed over a video-based training intervention, decreasing speed is best for decision-making development.
	Put, Wagemans, Spitz, Williams, and Helsen (2015)	1st person simulation – youth performers *and* computer animations from bird's-eye view	Explicit	8 weeks (1.5) - 360 videos & 360 animations	Intervention (10), control (10)	Training group improved in both the video and animation tests following the intervention. Control group did not change.	Combination of video-based training and computer animations is beneficial for assistant referees' offside decision-making.
	Put, Wagemans, Jaspers, and Helsen (2013b)	1st person simulation – youth performers *and* computer animations from bird's-eye view	Explicit	4 weeks (1) - 120 video & 120 animations	Intervention (10), control (8)	Training group improved following intervention in both the video and field test (simulated offside scenarios).	Video-based training can lead to an improvement in on-field and off-field decision-making.

Table 2.2 (continued)

Sport	Study	Video format	Instructional type	Training duration (no. sessions per week) - clips	Groups (participant no.)	Main effect	Conclusion
	Catteeuw, Gilis, Wagemans, and Helsen (2010c)	1st person simulation – youth performers *and* computer animations from bird's-eye view	Explicit	4 weeks (1) - 80 videos & 80 animations.	Training (10), Control (14)	Significant improvement training group pre to post. No change for control. Significant difference between groups.	Combination of video-based training and computer animations is beneficial for assistant referees' offside decision-making.
	Catteeuw et al. (2010b)	1st person simulation – youth performers *and* computer animations from bird's-eye view	Explicit	4 weeks (1) - 120 videos or 120 computer animations	Video training (9), animation training (9), control (22)	Significant different between groups following intervention ($p < 0.05$) as assessed through videos.	Both training groups improved decision-making accuracy in post-test, control group did not.
Rugby union							
	Mascarenhas et al. (2005c)	1st person game footage	Explicit	1 session - 25	Intervention (41), control (15)	Significant improvement following training only observed in the lower level officials ($p < 0.05$).	A short video-based training intervention appears to be more beneficial for lower level officials.

In terms of the duration of training, this is contingent on the instructional process used. A common instructional approach is explicit instruction, where participants are provided with cues or rules to follow (Gorman & Farrow, 2009; Mascarenhas et al., 2005c; Williams, Ward, & Chapman, 2003). Explicit instruction is an effective way to hasten decision-making development, particularly in amateur participants (Raab, 2003). Alternatively, implicit instruction uses minimal instructional cues or feedback in the training process (Raab, 2003). The implicit approach typically has greater skill retention and transfer to pressure situations, but requires a longer training duration than the explicit approach (Masters, 1992; Smeeton, Williams, Hodges, & Ward, 2005). Larkin et al. (2018b) used 1,040 clips over 12 weeks due to the implicit nature (i.e., no instruction or feedback), with decision-making changes only observed in the less experienced participants. This has led to more explicit (i.e., including feedback) approaches being used in a shorter intervention. For example, this approach was beneficial for soccer referees using 144 clips over 7 weeks (Schweizer et al., 2011). On the other hand, van Biemen et al. (2018) reported blurred footage is more beneficial than regular match broadcast video in a one-session intervention using 70 clips. The rationale for using blurred footage was to develop the officials' ability to identify key kinematic information, as most errors in the decision-making process are caused by missing information (MacMahon & Mildenhall, 2012). Finally, Kittel et al. (2019a) reported no improvements in an 8-week intervention with 96 clips for Australian football umpires. Given these mixed results of match broadcast video, additional video modes warrant further investigation.

As a ceiling effect may restrict potential improvements for more experienced and/or expert performers (Gorman & Farrow, 2009), the implicit approach may be more optimal with greater training time allocated (Larkin et al., 2018b). It is important to consider knowledge of the laws and decision-making are reported as being the two most important attributes for Australian football umpires (Kittel et al., 2019b). Explicit feedback may therefore be beneficial for this group given the importance of law knowledge when making decisions, as this produces more declarative knowledge through use of instructions and/or cues. The instructional technique for a video-based training intervention is contingent on the performance level of the participants and time available to elicit an effect.

Several studies have used video-based training for assistant soccer referees to develop offside decision-making skill, with positive effects (Catteeuw et al., 2010b; Put et al., 2013b; Put et al., 2015). These studies used computer animations, where officials identify the correct frame where a player has travelled offside. Animations are from a bird's-eye view, which may not be representative of the cues received in-game. These studies have also filmed first-person simulated offside scenarios with high performance youth players as actors. Researchers can carry out simulations with assistant referees as they are identifying spatial cues to determine whether a player is offside. No studies have simulated tackle/foul scenarios for soccer or Australian football officials due to the injury risk of players contacting each other forcefully. Although assistant referees are interactor officials, central referees in sports such as soccer, rugby union, and Australian football have more cues to monitor in their role and may require more complex training approaches (MacMahon et al., 2014).

2.3.5. Creating more representative video-based training tasks

To improve performance in these off-field video-based training tasks, it is imperative they represent specific constraints characteristic of the competition setting (Pinder et al., 2011). Therefore, researchers have adopted several approaches to increase the representativeness of video-based training by including constraints experienced in competition. One example of doing so is the above real-time training method, where footage is presented at faster speeds, where athletes make automatic decisions under significant time constraints (Lorains et al., 2013a). This form of video-based training enables participants to make more rapid, automatic decisions than they would in competition, leading to higher fidelity of the task (Lorains et al., 2013b). Above real-time training has resulted in decision-making improvements for Australian football players (Lorains et al., 2013a). Manipulating video speed can also be beneficial for assistant soccer referees, if scheduled in a "decreasing speed group" where video speed decreases from 125% to 100%, then 75% over the course of the intervention (Put et al., 2016). Whereas it appears beneficial for videos to be presented above real-time for athletes (Lorains et al., 2013a), the findings of Put et al. (2016) suggest a decreasing sequence of video speed (i.e., from faster to slower speeds over the intervention) is effective for officials.

Another example of including sport-specific constraints is incorporating physical fatigue, given the high physical loads officials experience within games (Elsworthy et al., 2014). Kittel et al. (2019a) incorporated video-based training into high intensity interval training, as officials make in-game decisions under high physical loads. However, these decisions were made stationary following a high intensity interval. Making decisions while moving would theoretically be more representative of in-game decision-making. The results of this study suggested combining video-based training with high intensity physical training is not beneficial for developing decision-making skill. This intervention did not significantly improve performance, suggesting more research is required to assess proposed representative interventions, despite the literature suggesting representativeness is key (Farrow et al., 2013, Hadlow et al., 2018).

2.3.6. Limitations of existing video-based training

There are several limitations of previous video-based training studies such as the use of match broadcast footage. For off-field tasks such as video-based training to represent the competitive environment, it is imperative these tasks maintain similar sources of information to those perceived in competition (Dicks et al., 2010). Match broadcast is the most common video mode in central officials such as Australian football field umpires (Kittel et al., 2019a; Larkin et al., 2018b) and soccer referees (Schweizer et al., 2011; van Biemen et al., 2018). As match broadcast footage is filmed from a fixed, elevated position in the grandstand, this presents a third-person perspective, dissimilar to the perspective an official would have in a game (Craig, 2013). To overcome this limitation, research has implemented first-person viewpoints to increase the representativeness of the task (Petit & Ripoll, 2008). Footage filmed from a first-person perspective would theoretically increase ecological validity as the perceptual information is more similar to in-game processes than an elevated, third-person perspective (Mascarenhas, Collins, & Mortimer, 2005a). Although first-person viewpoints increase the similarity of information received from on-field to video-based tasks, the vision does not change automatically with head movements, therefore limiting the behavioural correspondence of the task (Craig, 2013; Pinder et al., 2011). To increase the representative nature of the task, virtual reality technology has been

suggested as the next frontier for off-field perceptual-cognitive development (Farrow et al., 2018). This technology has the ability to overcome limitations of extant video-based training tools by including first-person viewpoints and updating vision from a participant's head movements.

2.4. Virtual reality

2.4.1. Defining virtual reality

Virtual reality is defined as simulations of a real or imaginary environment, where a participant can both perceive and interact with the environment (Craig, 2013; McMenemy & Ferguson, 2007). A key distinction of virtual reality is the concept of presence, where an individual feels they are within the virtual environment (Jensen & Konradsen, 2018; Steuer, 1992). The level of immersion experienced by an individual is dependent on the level of presence felt within the environment (Düking, Holmberg, & Sperlich, 2018). Immersion is a considerable factor in the degree of transferability from the off-field task to on-field environment (Brault, Kulpa, Duliscouët, Marin, & Bideau, 2015; Craig, 2013). The higher sense of immersion contributes to stronger task fidelity, which is a key component of representative design (Craig, 2013; Pinder et al., 2011). In essence, virtual reality is a broad term to describe technologies which enable a stronger sense of immersion and transfer to real-world tasks (Gray, 2017).

Virtual reality technology has developed over the years and can be presented in a number of formats, including flat or curved large screen displays, Cave Automatic Virtual Environment (CAVE) (where participants are in a room surrounded by a screen) and head mounted displays (HMD). Virtual reality has been used in a range of areas; including military (Reger et al., 2011), medical surgery (Gurusamy, Aggarwal, Palanivelu, & Davidson, 2009), and education settings (Freina & Ott, 2015). In sport, virtual reality has received a greater focus in recent years where it is commonly used for endurance sports to incorporate factors such as competitiveness, pacing and greater immersion (Neumann et al., 2018). For ball sports, virtual reality has been investigated with interceptive skills using different displays, including CAVE and HMD modalities (Faure, Limballe, Bideau, & Kulpa,

2020). Although CAVE systems are an effective presentation mode to increase levels of presence, they are very large and expensive to develop (Miles, Pop, Watt, Lawrence, & John, 2012).

2.4.2. Head mounted display technologies

For a more accessible virtual reality modality, HMDs are commonly used in sporting studies, where a key feature of this technology is stereovision/binocular vision (Faure et al., 2020). As outlined by Faure et al. (2020), the greater number of cues in stereovision contributes to stronger depth perception, presence (Craig, 2013), and embodiment (Kilteni, Groten, & Slater, 2012). Embodiment refers to the integration of different sensory signals, where the brain generates a self-representation (Matamala-Gomez et al., 2019). Embodiment is important for simulated practice by allowing the sensorimotor system to be more engaged, facilitating the development of behavioural changes (Bohil, Alicea, & Biocca, 2011). Previous two dimensional, screen-based approaches limit the stereoscopic information available to extract such as in real life (Vignais et al., 2015). HMDs can also lead to a stronger interaction with an environment than two-dimensional screen approaches, where vision automatically updates with head movements of the individual (Bird, 2020; Craig, 2013). The head movements when wearing a HMD contribute to stronger behavioural correspondence of the task, which is a key component of representative design (Pinder et al., 2011). Unlike existing screen-based approaches, the head movements experienced from using a HMD incorporates all components of the gaze control system; head, eyes, and body (Panchuk et al., 2018; Vickers, 2007). HMD technology has become a common method to present both virtual reality and 360° VR in a more immersive environment (Jensen & Konradsen, 2018).

2.4.3. Existing perceptual-motor and cognitive virtual reality investigations

Initial studies investigating virtual reality (i.e., virtual/animated footage) have been in endurance-based sports for pacing strategies (Neumann et al., 2018). Virtual reality research investigating perceptual-motor and perceptual-cognitive skills have begun to emerge. Studies have

investigated the use of virtual reality for skills such as intercepting the ball for goalkeepers in handball (Vignais et al., 2015). A comparison between virtual environments as viewed through a large cylindrical screen (i.e., CAVE environment) and standard video screen was conducted by Vignais et al. (2015). Handball goalkeepers made faster and more accurate decisions in the virtual reality condition than using video. Research has also examined the use of virtual reality for detecting deceptive body movements in rugby players, viewed through a HMD (Bideau et al., 2010; Brault et al., 2012). Validity was inferred in both studies, where experts outperformed novices in efficiently and accurately detecting deceptive movements (Bideau et al., 2010; Brault et al., 2012). Each of these studies used virtual characters as the visual stimulus for the virtual environment. There are no studies formally analysing the validity and reliability of novel virtual reality technologies for perceptual-cognitive skills.

There are a limited number of studies which have assessed the effectiveness of a virtual reality training intervention for sport-specific skills (Petri et al., 2018). Gray (2017) examined the efficacy of virtual reality technology to improve the perceptual-motor skill of baseball batting. Findings indicated a six-week virtual reality intervention led to an improvement in batting performance, with improvements retained one month following training (Gray, 2017). These results are analogous to a virtual reality intervention for table tennis players, which elicited a significant performance improvement following training (Michalski et al., 2019). The above studies mainly focused on perceptual-motor skill development, with an element of anticipation involved. Virtual reality implementing virtual environments often investigates perceptual-motor skills due to the high behavioural correspondence this technology enables (Hadlow et al., 2018). Düking et al. (2018) outlined a key strength of virtual reality whether viewed through a HMD, large screen or CAVE environment is the ability to couple motor actions in the virtual task. No studies to date have used virtual environments to improve the executive function of decision-making in sport.

A virtual reality environment has recently been developed as a potential training tool for soccer referees (Gulec, Yilmaz, Isler, O'Connor, & Clarke, 2019). Gulec et al. (2019) aimed to increase the exposure of referees to atmospheres they are likely to encounter in competitive games. The virtual environment affords manipulation and control of different competitive scenarios to expose officials to

these in a safe environment. Although this was not a training intervention, participants perceived the first-person viewpoint and game-likeness to be potentially suitable for use as a training tool. The physical fidelity may not be as strong for virtual environments such as these, as they do not present real world footage (Stoffregen et al., 2003). It may also be difficult to capture the kinematic information of dynamic tackling scenarios in Australian football (Larkin et al., 2018a) in virtual environments such as these. Therefore, researchers should explore approaches with stronger physical fidelity.

2.4.4. Virtual reality and 360° VR environment comparison

In terms of the type of footage used within virtual reality technology, there have been contrasting approaches to using either virtual/animated or real-world footage, where both have distinct advantages. The use of virtual scenarios has been commonly used as this provides greater control over the environment, manipulating the virtual environment and enables researchers to standardise conditions (Düking et al., 2018; Faure et al., 2020). Although virtual environments provide greater control for researchers and practitioners, visual correspondence is a key consideration of representative learning design (Pinder et al., 2011). A limitation of virtual environments is the lack of content or it is financially expensive and time consuming to develop such content (Düking et al., 2018). 360° VR footage viewed through a HMD is beginning to overcome this limitation where footage can be sourced in a more affordable manner (Jensen & Konradsen, 2018; Panchuk et al., 2018). A key distinction between virtual reality and 360° VR is that virtual reality enables interaction within the virtual environment, whereas 360° VR is video only. 360° VR may also lead to stronger ecological validity as the perceptual information is more similar to the competitive environment (Araujo et al., 2007) as it uses real world footage from a first-person viewpoint filmed using a 360° camera. Using real world 360° VR footage is a more visually realistic method, presenting a 'middle ground' between virtual environments and existing video-based (match broadcast perspective) approaches (Fadde & Zaichkowsky, 2018). This is evident in Table 2.3 which provides a brief summary differentiating between these technologies. 360° VR is a more middle ground as it creates more interaction for the perceptual, rather than the interactive/action component of the skill. The greater interaction of the

perceptual component allows for more affordances for action, where individuals can search to make a more informed decision as in a game. Theoretically, affordances for action are a key component of ecological dynamics which emphasises the importance of the relationship between perception, cognition and action (Renshaw et al., 2018). Furthermore, 360° VR is beneficial for perceptual-cognitive tasks where specific motor actions such as a pass or dodge are not fundamentally required (Fadde & Zaichkowsky, 2018). As officials do not require a motor action, but verbalise their decision, this may be an appropriate tool to use for this population, as this would be more representative of their in-game actions.

Table 2.3: Differentiation between match broadcast, 360° VR and virtual reality technologies.

Technology	***Type of footage***	***Perspective***	***Presentation method***	***Proposed benefits***	***Limitations***
Match broadcast video					
	Real world	Third-person	Screen (e.g., iPad, TV computer)	Presents proper game footage Easy to collect	Third-person perspective lacks ecological validity and representativeness Screen-based approaches limit representativeness as video does not update automatically with head movements
360° VR					
	Real world	First-person	Head mounted display	First-person perspective leads to stronger fidelity Suggested tool for 'read only' tasks where participants do not require a sporting action (e.g., officials) Stronger behavioural correspondence viewed with head movements	Limited interaction within environment Difficult to combine perception and action (such as a kick or a pass for athletes)
Virtual reality					
	Virtual (animated)	First-person	Head mounted display Large screen CAVE environment	Greater control over scenarios Combine perception and action	Presents animated video – difficult to capture dynamic sporting actions Financially expensive

2.4.5. Existing 360° VR training approaches

360° VR technology is a suitable method to train decision-making skill rather than perceptual-motor skills, as this technology is 'read only' where it does not allow similar interaction within the environment as virtual designs (Fadde & Zaichkowsky, 2018). Recently, studies have examined the efficacy of 360° VR as a tool to develop the perceptual-cognitive skill of decision-making in sport, specifically with basketballers. Both studies using this technology (Pagé et al., 2019; Panchuk et al., 2018) utilised small-sided game scenarios to film the immersive (360° VR) video. Panchuk et al. (2018) developed a 3-week training intervention for youth basketballers, with performance improvements following training assessed in a 360° VR post-test and small-sided game transfer test. There were large, non-significant improvements for the male intervention group in relation to the control group for both tests. Results for female participants in this study revealed both intervention and control female groups improving in the 360° VR post-test. The female control, rather than intervention group improved in the small-sided game transfer test. This was the first study to assess the effectiveness of 360° VR to improve sport decision-making skill, yet there were several limitations. These include only 41 clips used for the intervention, lack of a retention test, small group sample sizes, and gender of the players filmed.

Pagé et al. (2019) also developed a 360° VR training intervention for basketball players, with a further comparison made to two-dimensional screen-based video of the same clips. This study included three groups; 'VR' (i.e., 360° VR), CS (i.e., same footage as VR group but presented on a computer screen) and control. There were 200 clips observed overall for the intervention for each group, with 50 per each of the four training sessions. The control group watched footage from college basketball playoff games (i.e., match broadcast perspective), yet this footage was not structured as per the intervention groups, as it presented 15 minutes of continuous match play rather than isolated clips. Unlike Panchuk et al. (2018), decision-making changes were not assessed using a video-based (i.e., immersive 360° VR) test, but with an on-court transfer test. Researchers assessed decision-making though both trained and untrained plays (i.e., scenarios) in the on-court test. For trained plays, both intervention groups improved significantly compared to the control following the intervention. Improvements were only observed for the VR group in untrained plays, with no such changes for the

computer screen or control groups. This study provides promising findings for the use of 360° VR in developing decision-making skill in sport.

2.4.6. Research gaps for use of 360° VR technology

Results of the above studies are encouraging for the use of this technology in future decision-making training studies. There are several gaps in the literature for future studies to investigate. Video-based training programs, whether using existing match broadcast or novel 360° VR technologies need to establish the reliability and validity of the testing instruments (Larkin et al., 2015). If reliable and valid, performance in these video-based tests should be compared to in-game decision-making. To date, no studies have provided a matched comparison of the effectiveness of both 360° VR and previously used match broadcast footage on developing decision-making skill. A matched comparison would involve the same type and number of clips in both video conditions to allow consistency in analysis. Furthermore, no studies have examined the effectiveness of 360° VR technology in enhancing sports officials' decision-making skill.

2.4.7. Proposed benefits of using 360° VR

There are a number of proposed benefits for the use of virtual reality in sport, specifically for assessing and training perceptual-cognitive skills. The two studies using 360° VR for decision-making development (Pagé et al., 2019; Panchuk et al., 2018) both used the MPTF as the rational for using this technology (Hadlow et al., 2018). Virtual reality provides stronger behavioural correspondence, which can make a more representative task (Pinder et al., 2011). This can be achieved in simulations using a virtual environment as participants can interact, such as blocking a virtual ball (Vignais et al., 2015). 360° VR may be appropriate for sporting officials who perform perceptual-cognitive tasks, and their 'response' is verbal, rather than a sport-specific motor action. There remains an element of behavioural correspondence when using 360° VR due to the head movements to scan the 360° space while wearing the HMD. Unlike traditional screen-based approaches such as match broadcast footage where the

camera angle changes, vision automatically updates when wearing a HMD (Craig, 2013), leading to stronger behavioural correspondence.

As discussed throughout this review, fidelity is an important consideration to create tasks representative of the competitive environment (Farrow, 2013). Theoretically, 360° VR may provide stronger psychological fidelity and ecological validity than existing match broadcast perspectives, due to the first-person viewpoint which allows for more similar perceptual information to what is received in-game (Craig, 2013). In addition to psychological fidelity, 360° VR would theoretically have stronger physical fidelity than virtual environments due to the real world footage used, which would overcome a significant weakness of virtual reality approaches (Düking et al., 2018).

To summarise, 360° VR warrants further investigation as a tool to both assess and develop decision-making skill in sport. The 'read only' nature of 360° VR (Fadde & Zaichkowsky, 2018) could be considered a limitation for use in athletes, given the need to couple perception and action by executing a sport-specific movement (Craig, 2013). This does not pose a limitation for sporting officials such as Australian football umpires, given they do not complete a motor action when providing a perceptual response. Considering decision-making is the most important skill for sporting officials (Kittel et al., 2019b; Morris & O'Connor, 2016), the efficacy of technologies such as 360° VR should be examined for assessment and training purposes.

Chapter 3: Video-based Testing in Sporting Officials: A Systematic Review

This Chapter is presented in pre-publication format of a recent publication titled:

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

VICTORIA UNIVERSITY

THE NEW WAY TO DO UNI

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book:	**Kittel, A.**, Larkin, P., Elsworthy, N., & Spittle, M. (2019d). Video-based testing in sporting officials: A systematic review. *Psychology of Sport and Exercise*, 43, 261-270. https://doi.org/10.1016/j.psychsport.2019.03.013
Surname: Kittel	First name: Aden
Institute: Institute for Health and Sport	Candidate's Contribution (%): 75

Status:

Accepted and in press: Date:

Published: Date: July 2019

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be included in the thesis as outlined in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted] 12/8/2020

Signature **Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

PO Box 14428, Melbourne,
Vic 8001, Australia
+61 3 9919 6100

Victoria University ABN 83776954731
CRICOS Provider No. 00124K (Melbourne),
02475D (Sydney), RTO 3113

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location(s)**:

All electronic data will be stored on the Victoria University R Drive, which is a secure research storage drive.

Name(s) of Co-Author(s)	Contribution (%)	Nature of Contribution	Signature	Date
Paul Larkin	10	Reviewer, provided feedback, minor edits.	[redacted]	12/8/20
Nathan Elsworthy	7.5	Reviewer, provided feedback, minor edits.		12/8/20
Michael Spittle	7.5	Reviewer, provided feedback, minor edits.		12/8/20

Updated: September 2019

3.1. Abstract

Objectives: Decision-making is the most important skill for sporting officials, consequently, assessment of this skill is becoming increasingly popular in the literature. There is considerable interest in the use of video-based methods to assess decision-making of officials in controlled, off-field environments. *Design*: Systematic review of the literature examining video-based testing in sporting officials. *Methods*: Using the keywords "umpire", "referee", "sport officials", "decision making" and "judgement", a comprehensive search was conducted in February 2018 on electronic databases (SPORTDiscus, Medline, PsycInfo, Google Scholar). Inclusion criteria included full text articles from January 2000 to January 2018 published in peer-reviewed journals. Only 'central' or 'field' officials were included in this review (i.e., assistant referees, touch judges were excluded). *Results*: The search yielded 27 studies. The majority of articles were specific to soccer officials. Overall, video-based testing appears to be a valid measure of decision-making differentiating between performance levels. This review highlighted a high degree of variability among the methods applied, with varied participation groups, clip type used, and influences on decision-making. The reporting of reliability and implementation of transfer tests was rarely incorporated in the research. *Conclusions*: Video-based testing appears to be a valid measure of decision-making of officials in an off-field, controlled environment. This research area would be advanced through further investigation into sports other than soccer, examination of transfer to match performance testing, reporting the reliability of the test, reporting decisional accuracy rather than solely number of decisions, and investigation of additional video modes.

Keywords: Video-based testing, sports officials, perceptual-cognitive expertise.

3.2. Introduction

Perceptual-cognitive skills are an integral aspect of sporting performance for all individuals. Perceptual-cognitive skills are defined as the ability to identify crucial information within the environment, and integrate this information with existing knowledge of motor capabilities to select and execute an appropriate response (Marteniuk, 1976). Decision-making is the foremost perceptual-cognitive skill involved in sport (Williams, Ward, Smeeton, & Allen, 2004), and can be defined as the ability to perceive information, correctly interpret, then select an appropriate response (Baker, Côté, & Abernethy, 2003). The literature has focused on athletes, commonly investigating the development of decision-making (Baker, Cote, & Abernethy, 2003), expert-novice differences (Williams & Ericsson, 2005), and decision-making processes (Araújo, Davids, & Hristovski, 2006). Importantly, perceptual-cognitive skills are an effective means to differentiate between less skilled and higher skilled performers in sport (Berry, Abernethy, & Côté, 2008; Williams & Ericsson, 2005).

Sporting officials, however, are a vital component of the sporting domain along with players. As they are required to perceive sporting actions and react to whether an infringement (i.e., free kick, penalty) has occurred, decision-making is the foremost perceptual-cognitive skill of sporting officials. The accuracy of such decisions can greatly impact the outcome of a match, leading to criticism and impacting club revenue (Larkin, Berry, Dawson, & Lay, 2011). As such, decision-making is commonly cited as the most important overall skill for effective officiating (Helsen & Bultynck, 2004; Morris & O'Connor, 2016). Within the literature, officials are classified as interactors (e.g., soccer referee, Australian football umpire), monitors (e.g., volleyball referee; gymnastic judge), and reactors (e.g., tennis line judge) based upon their decision-making processes, and their interaction or movement within the environment (MacMahon et al., 2014). This review will focus on interactor officials who have high perceptual demands (i.e., cues and players to monitor), and interaction with their environment (high physical demands) (MacMahon et al., 2014).

From a sporting official's perspective, researchers have investigated both on-field and off-field aspects of performance. On-field research has typically examined the movement (Elsworthy &

Dascombe, 2011; Emmonds et al., 2015; Krustrup & Bangsbo, 2001) and decision-making (Burnett, Bishop, Ashford, Williams, & Kinrade, 2017; Helsen & Bultynck, 2004; Larkin, Mesagno, Berry, & Spittle, 2018) match demands of interactor officials in a range of sports. In relation to these areas of research, there are inconsistent findings between sports, with research suggesting there is no significant impact of exertion on match decision-making in Australian football umpires (Elsworthy, Burke, Scott, Stevens, & Dascombe, 2014) and rugby league referees (Emmonds et al., 2015), yet does impact the decision-making demands of soccer referees (Mallo, Frutos, Juárez, & Navarro, 2012). Further, the influence of communication (Cunningham, Simmons, Mascarenhas, & Redhead, 2014; Neville, Salmon, & Read, 2018), psychological factors (Johansen & Haugen, 2013; Page & Page, 2010) and physical fitness on match performance (Castagna, Abt, & D'ottavio, 2002) have also been investigated. These studies highlight that although there are similarities across different sporting officials, not all findings are transferrable across sports and therefore researchers should further investigate methods of assessing the decision-making performance of officials from a wide range of sports.

Off-field research in sporting officials has grown in recent years, with an emphasis on examining the perceptual-cognitive aspects of performance (i.e., decision-making) in an isolated manner. This is due to the paramount importance of decision-making to overall performance (Helsen & Bultynck, 2004). Assessing the decision-making of officials in games may be the optimum measure of decision-making performance, however, there is a high degree of variability from game to game. This limits the performance comparisons between officials across different games, and within individual officials from game to game. Off-field video-based testing overcomes this limitation, with the ability to test decision-making in a controlled environment to present consistent scenarios across multiple officials. We define these methods as presenting sport-specific decision-making in a video format to simulate on-field decision-making in an off-field setting. The high degree of variability in the methods applied for video-based research will be outlined as part of this review. Therefore, the aim of this systematic review is to summarise video-based decision-making assessment literature in the domain of interactor officials, and analyse the various methods utilised to simulate match-like decision-making.

3.3. Method

The method for this systematic review was informed by the PRISMA guidelines (Moher, Liberati, Tetzlaff, Altman, & Group, 2009), summarised in Figure 3.1.

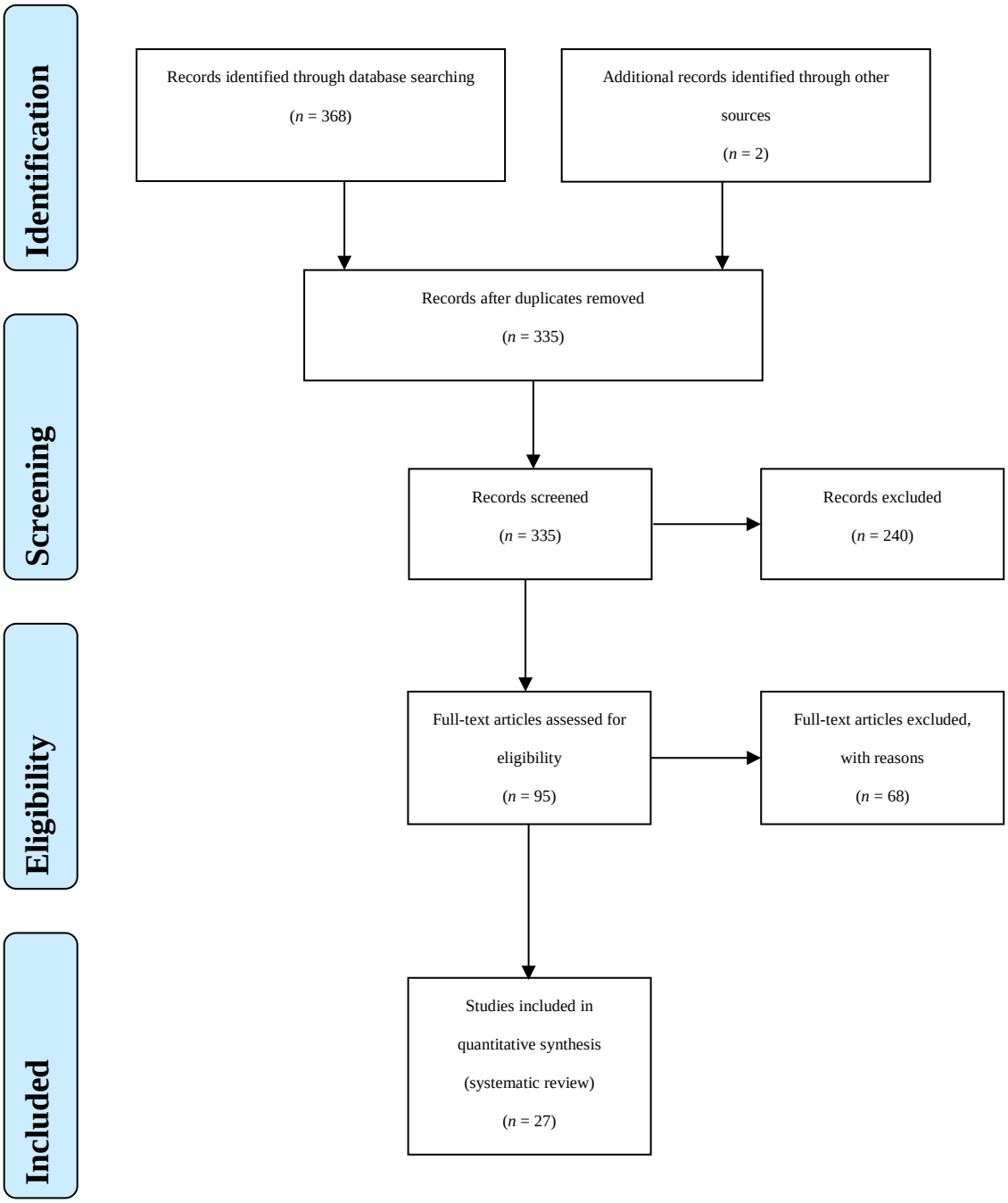

Figure 3.1: PRISMA flow diagram

3.3.1. Search strategies

Electronic databases (SPORTDiscus, Medline, PsycInfo, Google Scholar) were searched for articles published between January 2000 until February 2018. Keyword combinations included "decision making" in conjunction with "umpire" and "referee". The search was restricted to English peer-reviewed articles. 207 articles were identified in this initial search. Following this, keywords of the articles were analysed for further search combinations. Two new search terms were identified and combined with the existing terms including "judgement" and "sport officials", and subsequently searched within the databases. A further 161 articles were identified with the new search terms in the four databases, resulting in a total of 368.

3.3.2. Inclusion and exclusion of studies

Studies included in this systematic review adhered to the following criteria: (i) participant groups included interactor officials; (ii) used video-based methods for off-field decision-making assessment; (iii) assessed sport-specific decision-making involving infringement/penalty scenarios (i.e., general perceptual-motor skill assessments such as pattern recognition and reaction time were excluded); and (iv) participants were central referees/field umpires (i.e., assistant referees were excluded when the aim of the study investigated offside decision-making performance). These populations were excluded as this review focused on infringement-based decision-making, rather than onside/offside processing, which is subject to different psychological factors (e.g., flash-lag effect). Studies were included if infringement-based decision-making of central referees was compared to other populations, such as fans or assistant referees.

3.3.3. Screening articles

Each article was screened by examining the title, abstract and keywords based on the inclusion criteria. If there was any uncertainty over the appropriateness of an article, this was debated by the first

and second authors. In the rare circumstance where uncertainty remained, the third author was included in the discussion. Final classification and acceptance of all studies was agreed upon by all authors.

3.3.4. Quality assessment

The quality of articles was analysed using a scale adapted from previous research (Larkin, Mesagno, Spittle, & Berry, 2015). This assessment scale (Figure 3.2) assesses the quality of each study based on three sub-scales (assessment of test measures, groups examined, decision reporting). The quality of the test measures, including different levels of validity (face, construct and concurrent) and reliability of the test were assessed in the first sub-scale. Face validity was demonstrated if the test presented a sport-specific decision-making scenario (Larkin, Mesagno, Berry, & Spittle, 2014). In psychological tests, construct validity refers to the degree to which a test measures a concept or construct that it intends to measure (DeVellis, 2016; Haynes, Richard, & Kubany, 1995). In comparison, for performance assessments, construct validity is also obtained through evaluation of performance between known skill level differences (Gadotti, Vieira, & Magee, 2006; Thomas, Silverman, & Nelson, 2015). The known-group difference method (Thomas et al., 2015) is commonly used in performance-based tests, such as video-based tests, to determine construct validity of the measure (Larkin et al., 2015). For the purpose of this review, this is how construct validity will be defined. Concurrent validity refers to the relationship between the measure (i.e., decision-making performance) and a criterion such as performance ranking, or on-field performance (Gadotti et al., 2006). The second sub-scale examined the demographics of the participants, in relation to the level they typically officiate. This is an example of construct validity, and is important to determine whether the video-based testing tool distinguishes between performance levels. The third sub-scale refers to the reporting given to the decisions provided by the participants. Specifically, did the study report accuracy of the decisions? This is important, as although studies report the number of free kicks or penalties made, differences must be put into perspective by providing the accuracy of the decisions. The quality assessment of each study is presented in Table 3.2.

Strength of evidence	Assessment of Test Measure	Officiating Level Used	Clip Decision
Most Robust	Reliability and validity assessed	Three participating groups (elite, sub-elite, amateur)	Decision accuracy reported
	Reliability only		
	Concurrent validity only	Two participating groups only	
	Construct validity only	One participating group only	Only number of decisions reported
	Face validity only		
Least Robust	No reliability or validity assessed	Participating group not reported	No reporting of decision accuracy or number

Figure 3.2: Classification scale of the video-based testing tool strength of evidence.

Based on three independent variables [adapted from Larkin, Mesagno, Spittle, and Berry (2015)].

3.4. Results

Full-text review of 95 articles was conducted after being identified as potentially relevant from scrutinising titles, keywords and abstracts. To determine the appropriateness, the full text articles were reviewed and assessed against the aforementioned inclusion criteria. Reference lists of each article were also examined to include articles not located within the above search criteria, with two studies included. Studies were predominantly excluded because they examined on-field decision-making (n = 56), rather than off-field video-based assessment per the criteria. Overall, 27 studies were included in the final analysis.

Table 3.1: Summary of studies examining video-based decision-making in sporting officials.

Author (date)	Sport officials' skill level	Other populations	Test overview	Influences on DM in task	Validity and Reliability Assessment	Results overview	Key findings
Soccer							
Plessner and Betsch (2001)	Skill level NR (n = 58)	Players (n = 57)	20 match clips.	Impact of previous decisions.	NR (not reported or evident)	No significant differences between referees' and players' decision-making scores. Referees did not want to award a penalty in the second scene if they had awarded a penalty in the first.	Both referees and players were biased by their own earlier decisions.
Jones, Paull, and Erskine (2002)	Amateur (n = 38)	Nil.	50 match clips (20 'certain' where participants called a foul 90% of the time, 20 'uncertain' where a foul was called 45%, 10 'innocuous' where foul committed 11%).	Aggressive reputation. Participants told those wearing blue uniforms had a history of aggressive behaviour (experimental group), or not told (control).	NR	No significant difference in number of decisions awarded against the blue team in either conditions ($p > 0.01$). This was evident for both 'certain' and 'uncertain' clips. Significant difference ($p < 0.01$) by number of cards by the two groups.	Referees who are informed of a team's aggressive reputation respond differently than those who do not receive this information, by awarding more red and yellow cards.
Nevill, Balmer, and Williams (2002)	Amateur (n = 40)	Nil.	47 match clips (22 referees watched the clips with noise, 18 in silence)	Crowd noise	NR	Referees in the noise group awarded significantly less fouls against the home team than the silent group ($p < 0.05$). More experienced referees awarded significantly less fouls ($p < 0.05$).	Crowd noise influences decision-making, by reducing the number of fouls against the home team rather than increasing fouls against away.
Balmer, Nevill, Lane, and Ward (2007)	Individuals with coaching, playing and/or refereeing experience (n = 26)	Nil.	One game including 47 incidents. Participants randomly assigned to a silent, and a noise group (with commentary). Tested on opposite condition one week later.	Crowd influence, anxiety levels.	NR	In noise condition, referees awarded fewer fouls against the home side ($p < 0.001$), and more no foul decisions ($p = 0.003$). Significant relationship between noise condition, and cognitive anxiety ($r = 0.55$, $p < 0.01$), and mental effort ($r = 0.54$, $p < 0.05$).	Supports the home crowd advantage theory. Biased decision-making was linked to increase in cognitive anxiety and mental effort.

Table 3.1 (continued)

Author (date)	Sport officials' skill level	Other populations	Test overview	Influences on DM in task	Validity and Reliability Assessment	Results overview	Key findings
MacMahon, Helsen, Starkes, and Weston (2007a)	Elite (n = 7)	Youth academy players (n = 41)	20 clips.	Nil	Construct validity only.	Referees scored significantly higher (80.6%) than players (55.1%) ($p < 0.001$). No significant effect of playing experience on decision-making accuracy ($p > 0.05$). Hours per week of practice is a moderate predictor of accuracy ($p < 0.01$).	Soccer referees significantly outperform players in an infringement identification task. Practice activities predict performance.
Catteeuw, Helsen, Gilis, and Wagemans (2009)	Elite (n = 27)	Elite assistant referees (n = 27)	24 clips presented twice. First as normal speed. Second set as once normal speed, followed by the same clip twice in slow motion.	Nil.	Construct validity only.	Referees scored significantly higher (72.4%) than assistant referees (66.9%) ($p < 0.01$).	Referees outperform assistant referees in an infringement identification task.
Ghasemi, Momeni, Jafarzadehpur, Rezaee, and Taheri (2011)	Elite (n = 41). Split into two groups- top 10 ranked, bottom 10 ranked	Nil.	76 match clips	Nil	NR	Top referees scored significantly higher than bottom ranked.	This video test was able to distinguish between competition levels for decision-making skill.
Poolton, Siu, and Masters (2011)	Amateur (n = 28). Group split in half according to whether they were high or low ruminators.	Nil.	45 match clips.	Effect of previous decisions	NR	The high decision rumination group awarded significantly more fouls against the away team ($p = 0.001$).	Ruminating over previous decisions can have an influence on awarding penalties to the home team.
Wagner-Egger, Gygax, and Ribordy (2012)	Amateur (n = 17)	Amateur players (n = 43) Fans (n = 22)	64 clips of the video game FIFA 2005.	Racism.	NR	Generally, referees were more likely to evaluate challenges as fouls than players and fans ($p < 0.05$). Referees were more likely to evaluate challenges by white players as fouls, than black players ($p < 0.05$). This was more evident in the referee group than players and fans.	There is some evidence of discrimination, but not all in black players.

Table 3.1 (continued)

Author (date)	Sport officials' skill level	Other populations	Test overview	Influences on DM in task	Validity and Reliability Assessment	Results overview	Key findings
Krenn (2014)	Sub-elite (n = 42)	University students with high football law knowledge (n = 81) University students with low football law knowledge (n = 82)	52 match clips (24 chromatic, 28 achromatic)	Uniform colour	NR	Referees judged tackles less harshly than university students with a high and low knowledge of football ($p < 0.01$). Colour did not have a significant impact on tackle judgements ($p > 0.05$).	Uniform colour does not affect judgement of tackles.
Renden, Kerstens, Oudejans, and Cañal-Bruland (2014)	Sub-elite (n = 31)	Sub-elite players (n = 17) Wheelchair-bound fans (n = 12) Novices (n = 18)	54 clips	Previous motor experience	Construct validity only.	Players and referees were both significantly more accurate than fans ($p < 0.001$). No significant differences between players and referees ($p < 0.998$), nor fans and novices ($p < 0.799$).	Motor experience playing the sport (i.e., players), is beneficial for decision-making accuracy compared to no motor experience (wheelchair bound fans). Players and referees have similar decision-making skill.
Lex, Pizzera, Kurtes, and Schack (2015)	Amateur (n = 50)	Nil.	52 clips. Participants assigned to two groups based on age and refereeing experience. One group first watched a silent, and then a noise condition (with match noise). Tested on opposite condition one week later.	Crowd influence.	NR	Does not measure decision accuracy. No influence of sound on foul decisions (p = 0.806). Significant effect of sound on penalty (i.e., yellow or red card) ($p < 0.01$). Referees were significantly more likely to produce a yellow card when players produced audible vocalisations (30.8%) than no vocalisations (23.6%).	When the foul has already been made, players' vocalisations influence whether a yellow card will be produced.
Spitz, Put, Wagemans, Williams, and Helsen (2016)	Elite (n = 20) Sub-elite (n = 19)	Nil.	20 simulation clips from perspective of assistant referee	Nil	Construct validity only.	Participants significantly more accurate in corner kick than open play situations ($p <$ 0.001). Elite referees significantly more accurate in certain situations.	This video test was able to distinguish between competition levels for decision-making skill.

Table 3.1 (continued)

Author (date)	Sport officials' skill level	Other populations	Test overview	Influences on DM in task	Validity and Reliability Assessment	Results overview	Key findings
Nevill, Hemingway, Greaves, Dallaway, and Devonport (2017)	Sub-elite (n = 6)	Nil.	One game. 2 referees watched game with no supporters present, 2 watched with one team's supporters present, other pair watched with other team's supporters.	Crowd influence.	NR	The referees with NO supporters present were significantly more likely to disagree with match referee's decision (p = 0.004).	Systematic tendency of crowd to influence referees to make less decisions. Evidence of home bias present with referees tending to favour team of supporters present.
Spitz, Put, Wagemans, Williams, and Helsen (2017)	Elite (n = 19) Sub-elite (n = 18)	Elite assistant referees (n = 24)	40 clips presented twice (once in normal speed, once in slow motion. Order randomised).	Slow motion speed.	Construct validity. Reliability assessed by intra-class correlation coefficients over two viewings, high reliability (range 0.76–0.82).	Elite scored higher (66.4%) than sub-elite (60.1%), and assistant referees (58.0%). Both the sub-elite and assistant referees scored significantly lower than elite (p < 0.001). All referees were significantly more accurate in slow motion (p < 0.001).	Referees are more accurate in slow motion than real time. This video test was able to distinguish between population groups.
Spitz, Put, Wagemans, Williams, and Helsen (2018)	Elite (n = 22) Sub-elite (n = 21)	Nil.	20 simulation clips from perspective of assistant referee.	Nil	Construct validity. Test-retest reliability from previous study.	Elite referees were significantly more accurate than sub-elite (p = 0.012).	This video test was able to distinguish between competition levels for decision-making skill.
Australian football							
Larkin, Berry, Dawson, and Lay (2011)	Elite (n = 15) Sub-elite (n = 23)	Nil.	25 match clips.	Nil	Construct validity only.	Elite umpires scored significantly higher than sub-elite umpires (p < 0.05).	Elite umpires have more advanced decision-making skill than their sub-elite counterparts.
Larkin et al. (2014b)	Amateur (n = 15)	Nil.	32 match clips (8 following the completion of each game quarter)	Game physical exertion levels.	Construct validity. Reliability from Larkin et al. (2014a).	No significant correlations between physical exertion and decision-making in a particular quarter. Significant improvement in quarter 4 (p = 0.001) compared to quarters 2 & 3.	No relationship between decision-making and in-game physical exertion. Higher decision-making at the end of the game could be due to the high importance of this period.

Table 3.1 (continued)

Author (date)	Sport officials' skill level	Other populations	Test overview	Influences on DM in task	Validity and Reliability Assessment	Results overview	Key findings
Paradis, Larkin, and O'Connor (2016)	Sub-elite (*n* = 10) Junior (*n* = 8)	Nil.	50 clips (10 blocks of 5: 2 x "easy", 1 x "medium", 2 x "hard")	Physical exertion: 10 x 300m run test.	Construct validity. Reliability from Larkin et al. (2014a).	Sub-elite outperformed junior in decision-making ($p = 0.016$, $r = 0.5$). Only scored significantly higher in "easy" clips ($p = 0.043$, $r = 0.46$). No significant correlation between physical exertion and decision-making overall, and in any difficulty.	Physical exertion does not influence decision-making in an off-field decision-making test. Sub-elite umpires have better decision-making than junior.
Rugby union							
MacMahon and Ste-Marie (2002)	Study 1: high experience (*n* = 12), low experience (*n* = 12). Study 2: high experience (*n* = 12)	Study 2: players (*n* = 12)	Clip *n* NR	Nil	NR	High and low experience had the same accuracy (53.7%) in Study 1. Players (62%) had slightly higher accuracy than referees (59%) in Study 2. Statistical significance NR.	No significant differences in accuracy for either study.
Mascarenhas, Collins, and Mortimer (2005)	Elite (*n* = 45)	Referee assessors (*n* = 27) Referee coaches (*n* = 13) Touch judges (*n* = 47)	10 match clips. Referees split into three groups based on ranking.	Nil	Construct validity only.	Top group of referees scored highest. Bottom ranked referees scored higher than the middle group.	Low level of accuracy and agreement of decisions.
Nazarudin et al. (2015)	Rugby referees (*n* = 132. Experience: 1-5yrs *n* = 33, 6 – 10yrs *n* = 34, 11 – 15yrs *n* = 34, 16+yrs *n* = 31)	Nil.	18 match clips (filmed from referees' perspective)	Nil	Concurrent validity.	Highly experienced referees scored higher in all decision-making than less experienced ($p < 0.05$). Match performance measured; significant relationship between match performance and decision-making performance ($r = 0.61$, $p < 0.05$)	This first-person video test was able to predict match performance, and differentiate between experience levels.
Basketball							
Brand, Schmidt, and Schneeloch (2006)	Elite (*n* = 113)	Nil.	18 match clips. One condition in game/block order. One condition in random order.	Impact of previous decisions.	NR	Referees in the random condition more likely to make more rigorous decisions than original sequence decision.	Referees are less rigorous in their decision-making when presented in game order than random.

Table 3.1 (continued)

Author (date)	Sport officials' skill level	Other populations	Test overview	Influences on DM in task	Validity and Reliability Assessment	Results overview	Key findings
MacMahon, Starkes, and Deakin (2007b)	Amateur (*n* = 44)	Nil.	44 clips (2 sets of 22). Knowledge primed 1 (KP1): watched first set of infractions (ID1), then rules/signal test (T), then second set of infractions (ID2). IP1: watched ID2, then 1D1, then did T. KP2: watched ID2, then did T, then watched ID1. IP2: watched ID1, then ID2, then did T.	Priming before decision-making.	NR	Participants were more accurate in the set that had fewer off the ball infractions. Knowledge and infraction priming had little to no effect on decision-making.	More off the ball incidents represent a higher difficulty of decision-making task.
Handball							
Souchon, Coulomb-Cabagno, Traclet, and Rascle (2004)	Amateur (*n* = 30)	Nil.	Watched video clips (*n* = unknown) of male and female handball incidents. Equal number of penalties for each.	Decision-making differences of male and female sport.	NR	Significantly more penalties for women than men ($p < 0.05$). Women received significantly more disciplinary penalties ($p < 0.001$).	Referees perceive fouls and penalties differently between men & women.
Souchon, Livingstone, and Maio (2013)	Elite (*n* = 47) Sub-elite (*n* = 48) Amateur (*n* = 50)	Nil.	122 match clips (60 male, 62 female games)	Player gender.	NR	All participants more likely to apply sanctions to, and intervene in situations with female players than male players ($p < 0.001$), this was more evident in amateur/junior referees ($p < 0.001$). Non-significant difference between sub-elite and elite referees.	More penalties against female players than male players. Does not provide decision accuracy.
Ice hockey							
Wilson and Mock (2013)	Ice hockey (high certification *n* = 15, low certification *n* = 15). High certification was level 3 or higher.	Nil.	10 match clips. Seven (high accuracy was 5-7 correct calls) were a penalty, three no penalty (high accuracy was 3 correct calls.)	Nil	NR	Neither certification nor assertiveness levels were significantly associated with making correct calls, whether that be penalty or no penalty ($p > 0.05$). Highly certified referees that had higher assertiveness were more likely to make a correct call, referees with high certification and low assertiveness had lowest decision-making ($p = 0.03$)	Assertiveness levels are more strongly associated with decision-making in ice hockey referees rather than certification level

3.4.1. Sports investigated

Key results from each study are presented in Table 3.1. In the studies identified, the most prevalent sport investigated was soccer (*n* = 16), followed by Australian football umpires (*n* = 3) and rugby union officials (*n* = 3), handball referees (*n* = 2), basketball referees (*n* = 2), and ice hockey referees (*n* = 1).

3.4.2. Reliability and validity of tests

Reliability of the decision-making test was reported in four of the 27 studies. One study conducted a reliability assessment, with intra-class correlation coefficients demonstrating high test-retest reliability (ICC = 0.76 - 0.82) (Spitz, Put, Wagemans, Williams, & Helsen, 2017). Three studies reported reliability from previous studies (Larkin et al., 2014b; Paradis, Larkin, & O'Connor, 2016; Spitz, Put, Wagemans, Williams, & Helsen, 2018).

The three types of validity examined were face/content, construct, and concurrent validity. As the inclusion criteria stipulates video clips must include sport-specific decision-making for officials (i.e., presenting an infringement, penalty, or free kick), all the studies had adequate face validity. Although this was not explicitly reported by the researchers, face validity is evident due to the sport-specific decision-making nature of the task. Similarly, construct validity is not explicitly stated by the studies, but can be assumed with a comparison of different performance levels. Construct validity was assessed by differentiating between known skill levels in 13 studies (Catteeuw, Helsen, Gilis, & Wagemans, 2009; Larkin et al., 2011; Larkin et al., 2014b; MacMahon, Helsen, Starkes, & Weston, 2007a; MacMahon & Ste-Marie, 2002; Mascarenhas, Collins, & Mortimer, 2005a; Paradis et al., 2016; Plessner & Betsch, 2001; Renden, Kerstens, Oudejans, & Cañal-Bruland, 2014; Spitz, Put, Wagemans, Williams, & Helsen, 2016; Spitz et al., 2017; Spitz et al., 2018; Wilson & Mock, 2013). Of the 13 studies which assessed construct validity, no significant differences were found in three of the studies (MacMahon & Ste-Marie, 2002; Plessner & Betsch, 2001; Wilson & Mock, 2013). This infers the decision-making task was not able to differentiate decision-making skill between known performance

levels. Concurrent validity, which examines the correlation between test score and on-field performance or ranking, was assessed in one study (Nazarudin et al. 2015).

3.4.3. Skill levels

There was a cross-sectional analysis of one skill group investigated in 8 studies. These studies covered a range of skill levels, including amateur officials (n = 6) (Jones, Paull, & Erskine, 2002; Larkin et al., 2014b; Lex, Pizzera, Kurtes, & Schack, 2015; MacMahon, Starkes, & Deakin, 2007b; Nevill, Balmer, & Williams, 2002; Souchon, Coulomb-Cabagno, Traclet, & Rascle, 2004). Sub-elite officials alone were investigated in one study (Nevill, Hemingway, Greaves, Dallaway, & Devonport, 2017), and one study with only elite officials (Brand, Schmidt, & Schneeloch, 2006). There were ten studies identified which compared multiple skill levels of sporting officials. Specifically, two studies split one skill level into two participant groups based on the following criteria; high or low ruminators (i.e., considering previous decisions) (Poolton, Siu, & Masters, 2011), and top compared to bottom ranked (Ghasemi, Momeni, Jafarzadehpur, Rezaee, & Taheri, 2011). Two studies investigated the effect of experience (MacMahon & Ste-Marie, 2002; Nazarudin et al., 2015), and one investigated certification level (Wilson & Mock, 2013). In terms of officiating level; amateur, sub-elite and elite were compared in one study (Souchon, Livingstone, & Maio, 2013), elite vs sub-elite in three studies (Larkin et al., 2011; Spitz et al., 2016; Spitz et al., 2018), and sub-elite to junior in one study (Paradis et al., 2016). The differences between each of these groups are presented in Table 3.2.

The remaining studies (n = 9) investigated decision-making differences of interactor officials to individuals who do not complete 'central' officiating tasks. Comparisons were made against assistant referees (Catteeuw et al., 2009; Spitz et al., 2017), soccer players (MacMahon et al., 2007a; Plessner & Betsch, 2001), wheelchair-bound fans, players and novices (Renden et al., 2014), players and fans (Wagner-Egger, Gygax, & Ribordy, 2012), official's assessors, official's coaches, touch judges (Mascarenhas et al., 2005a), university students with high football knowledge and low football

knowledge (Krenn, 2014), and finally individuals with soccer refereeing, coaching and playing experience but did not distinguish between groups (Balmer, Nevill, Lane, & Ward, 2007).

3.4.4. Clip type used

There was a range of decision-making footage in the testing protocols. Most commonly, individual video clips of match play from a broadcast perspective (i.e., third-person) were presented in 21 studies (Brand et al., 2006; Catteeuw et al., 2009; Ghasemi et al., 2011; Jones et al., 2002; Krenn, 2014; Larkin et al., 2011; Larkin et al., 2014b; Lex et al., 2015; MacMahon et al., 2007a; MacMahon et al., 2007b; MacMahon & Ste-Marie, 2002; Mascarenhas et al., 2005a; Nevill et al., 2002; Paradis et al., 2016; Plessner & Betsch, 2001; Poolton et al., 2011; Renden et al., 2014; Souchon et al., 2004; Souchon et al., 2013; Spitz et al., 2017; Wilson & Mock, 2013). Two studies investigated the decision-making of referees while watching one soccer game from a broadcast (i.e., third-person) perspective in a single sitting (i.e., watching a full football game at once) (Balmer et al., 2007; Nevill et al., 2017). One study investigated individual clips filmed from the referee's perspective (i.e., first-person) using head-mounted glasses (Nazarudin et al., 2015). Two studies filmed simulation clips of players from the assistant referee's perspective (mix of first and third-person) attempting to present a stronger simulation of the referee's perspective (Spitz et al., 2016; Spitz et al., 2018). Although this is from the assistant referees' perspective, the decision-making was infringement-based as per the inclusion criteria. Clips from the video game FIFA 2005 were investigated in one study (Wagner-Egger et al., 2012).

3.4.5. Additional influences on decision-making

Two studies examined the effect of physical exertion on decision-making performance, with one implementing decision-making following 300m efforts (Paradis et al., 2016), and another in the quarter breaks of an Australian football match (Larkin et al., 2014b). Four studies examined the influence of crowd noise on decision-making performance. Of these, one study watched a game with a team's supporters present in the room (Nevill et al., 2017), and three studies implemented crowd noise

into the video clips compared to a silent condition (Balmer et al., 2007; Lex et al., 2015; Nevill et al., 2002). The effect of previous decisions was investigated in three studies (Brand et al., 2006; Plessner & Betsch, 2001; Poolton et al., 2011). One study explored the effect of priming prior to a decision-making test (MacMahon et al., 2007b). Two examples of priming were used, including; knowledge priming (i.e., completing a rules and signals test prior to the decision-making task), and infraction priming (i.e., being instructed to focus on specific infractions prior to the decision-making task). The influence of slow motion on decision-making accuracy was assessed in one study (Spitz et al., 2017). Biases were examined as an influence on decision-making in five studies encompassing player gender (Souchon et al., 2004; Souchon et al., 2013), player skin colour (Wagner-Egger et al., 2012), aggressive team reputation (Jones et al., 2002), and uniform colour (Krenn, 2014). There were 11 studies which did not investigate any additional influences on decision-making (Catteeuw et al., 2009; Ghasemi et al., 2011; Larkin et al., 2011; MacMahon et al., 2007a; MacMahon & Ste-Marie, 2002; Mascarenhas et al., 2005a; Nazarudin et al., 2015; Renden et al., 2014; Spitz et al., 2016; Spitz et al., 2018; Wilson & Mock, 2013).

3.4.6. Transfer of skills to match performance

One study examined the transfer of video-based performance to match performance (Nazarudin et al. 2015).

Table 3.2: Quality assessment of included studies.

	Validity and reliability			*Skill level of officials*				*Decision reporting*		
Study	**Not reported or evident**	**Construct or concurrent validity evident**	**Reliability reported**	**Not reported**	**Amateur or junior**	**Sub-elite**	**Elite**	**Not reported**	**Number of decisions only**	**Accuracy of decisions reported**
Spitz et al. (2018)		✓	✓			✓	✓			✓
Spitz et al. (2017)		✓	✓			✓	✓			✓
Paradis et al. (2016)		✓	✓		✓	✓				✓
Larkin et al. (2014b)		✓	✓		✓					✓
Spitz et al. (2016)		✓				✓	✓			✓
Larkin et al. (2011)		✓				✓	✓			✓
Catteeuw et al. (2009)		✓					✓			✓
MacMahon et al. (2007a)		✓					✓			✓
Nazarudin et al. (2015)		✓				✓				✓
Mascarenhas et al. (2005)		✓					✓			✓
Renden et al. (2014)		✓				✓				✓
Ghasemi et al. (2011)	✓						✓			✓
MacMahon et al. (2007b)	✓				✓					✓
Wilson and Mock (2013)	✓			✓						✓
MacMahon and Ste-Marie (2002)	✓			✓						✓
Plessner and Betsch (2001)	✓			✓						✓
Souchon et al. (2013)	✓				✓	✓	✓		✓	
Brand et al. (2006)	✓						✓		✓	
Nevill et al. (2017)	✓					✓			✓	
Krenn (2014)	✓					✓			✓	
Lex et al. (2015)	✓				✓				✓	
Wagner-Egger et al. (2012)	✓				✓				✓	
Poolton et al. (2011)	✓				✓				✓	
Souchon et al. (2004)	✓				✓				✓	
Nevill et al. (2002)	✓				✓				✓	
Jones et al. (2002)	✓				✓				✓	
Balmer et al. (2007)	✓			✓					✓	

3.5. Discussion

The primary aim of this review was to provide a summary of the off-field video-based decision-making assessment literature of interactor officials, and analyse the various methods utilised to simulate sport-specific decision-making. The results highlight several key findings, including: i) soccer (football) is emphatically the most researched sport in this domain; ii) off-field video-based methods appear to have high construct validity (i.e., ability to differentiate between skill levels); iii) a high degree of variability in the methods applied to each study, leading to mixed interpretations of the results; iv) most influences on decision-making applied in the tests effectively highlight potential biases that may be present to an official in match play. Importantly, tests that effectively replicate potential biases such as crowd noise present in match play may be able to delineate which officials are affected by certain biases, leading to individualised training programs to promote consistent decision-making.

This review focuses on the off-field decision-making assessment of interactor officials from a range of sports. Soccer officials are the most commonly researched group, followed by Australian football umpires and rugby union referees with three studies focusing on each of these groups. A review of research on sporting officials also indicated the majority focuses on soccer officials (Hancock, Rix-Lièvre, & Côté, 2015). While certain findings can be transferred from research involving soccer referees, there are a number of inherent differences between soccer and other sports. As noted by MacMahon et al. (2014), sporting officials in different sports vary in the number of players and cues they need to monitor as part of their role, which can consequently impact the decision-making requirement of officials across different sports. Furthermore, it has been suggested that a wider range of research is required to determine which officiating approach is most effective given a specific sport or situation (MacMahon et al., 2014). Different approaches to decision-making include officiating in a black and white manner where each decision is taken in isolation, or using contextual judgement to apply the laws in consideration of the environment. Specifically, a key coaching instruction of soccer referees is for each decision to be evaluated in isolation, irrespective of previous decisions (Plessner & Betsch, 2001). Despite these coaching instructions, there are certain number of "unwritten rules" in officiating which can influence decisions (Plessner & Betsch, 2001). Contextual judgement is regarded

as an important factor by elite officials themselves (Mascarenhas, Collins, & Mortimer, 2005b; Morris & O'Connor, 2016), whereby decision-making is not black and white, rather officials are somewhat flexible in their final decision based upon certain contextual factors. For example, elite rugby union and rugby league referees consider the impact of a number of factors such as time, score line, momentum and field position when making decisions (Mascarenhas et al., 2005b; Morris & O'Connor, 2016). This knowledge will contribute to the understanding of how officials approach decision-making, by further exploring the effect of contextual judgements and "unwritten rules" in specific sports.

A limited number of studies reported the validity of the video-based test, and fewer reported reliability. It is imperative measures of validity and reliability are provided prior to using a video-based assessment to ensure accurate results (Larkin et al., 2015), and is a necessity to ensure a robust measurement tool for accurate results. Many studies do not explicitly state the validity of their testing measures. Construct validity is, in fact, measured in 13 of the 27 studies included, by comparing decision-making performance across multiple skill groups. Construct validity was evident in 10 of these studies, which were able to differentiate between skill levels using the video-based task. In terms of reliability, this was reported in only four of the 27 studies. One study included a reliability assessment as part of the study, demonstrating high reliability as assessed per intra-class correlation coefficient (0.76-0.82) (Spitz et al., 2017). In addition, three studies reported reliability from a previous paper (Larkin et al., 2014b; Paradis et al., 2016; Spitz et al., 2017). Only one study examined concurrent validity of the decision-making task (Nazarudin et al. 2015). A video-based test with a strong relationship to on-field performance or ranking has the potential to be used in conjunction with a battery of additional performance measures for talent identification. These results highlight reliability is a key consideration to be included in future studies, yet is not in the majority of research on this topic.

A key measure of validity is construct validity, which in this review refers to how the test differentiates between known skill levels (Gadotti et al., 2006). It is evident there is a high degree of inconsistency within this research area, when examining a range of skill levels. Eight studies provided a cross-sectional analysis of just one skill level. Some of these studies investigated the effect of an external influence on decision-making performance such as crowd (Nevill et al., 2017) or player

vocalisation (Lex et al., 2015), yet is it important to consider whether different levels of sporting officials are affected by these external stimuli similarly or differently. Ten studies investigated the decision-making performance across skill levels, of which one study examined three separate performance levels; elite, sub-elite, and novice (Souchon et al., 2013). It is imperative for studies to draw comparisons across multiple skill levels to determine construct validity. Establishing the validity of a video-based tool assists in monitoring individual decision-making progression over time using a consistent measurement. When determining what constitutes the performance level of an official, there is a lack of uniformity among the studies, similar to the general sporting population (Swann, Moran, & Piggott, 2015). The majority of studies consider 'elite' officials to be officiating the top national level for their respective sports (Brand et al., 2006; Larkin et al., 2011; Souchon et al., 2013; Spitz et al., 2016; Spitz et al., 2018). This definition is accurate for Australian football umpires as Australia is the only country to play this sport at a national high performance level. However, this can become problematic when considering sports such as soccer and basketball. National German basketball referees are labelled 'elite' by Brand et al. (2006), yet the highest performance level for basketball would be the Basketball World Championships or the NBA competition in the United States. To assist with the creation of a more uniform system for comparison between studies, we propose a categorisation system based from Swann et al. (2015). This system provides examples from the three most researched sporting official groups in this paper (soccer, Australian football, rugby union) (see Figure 3.3). As noted by Swann et al. (2015), classification of officials is dependent on not only performance level, but also the standard of competition in the country the sport is officiated. For multi-national sports such as soccer and rugby union, national-level officials are classified differently in large or small sporting nations. It is anticipated the development of this continuum will assist the uniformity of performance levels examined within, and between sports in this research area.

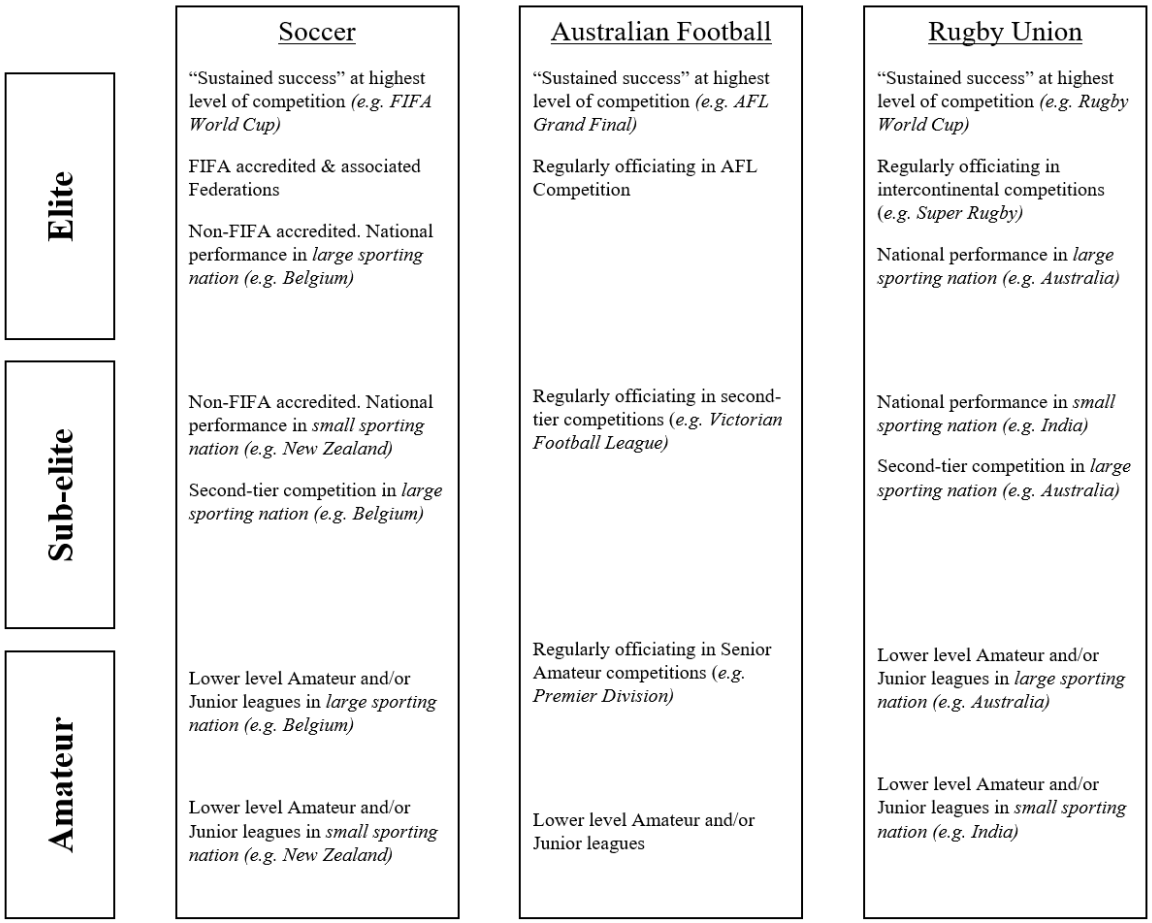

Figure 3.3: Classification of the different levels of sporting officials.

A number of studies compare decision-making performance of referees to other population groups. For example, foul infringement decision-making was assessed between central soccer referees, and their assistant referee counterparts (Catteeuw et al., 2009; Spitz et al., 2017). These studies were included as they assessed match-specific decision-making of the central referee, compared to the assistant referees completing the same task (i.e., no off-side decision-making was included). In both studies, the central referees scored higher than assistant referees. When comparing soccer referees to players, however, referees outperformed soccer players in decision-making performance in one study (MacMahon et al., 2007a), but not another (Plessner & Betsch, 2001). Similarly, Renden et al. (2014) determined there were no significant differences between soccer players and referees, yet higher decision-making performance by these groups than wheelchair-bound fans and novices (with no soccer experience). They concluded sport-specific motor experience may play an important role in decision-

making skill. Researchers are encouraged to assess accuracy of decisions, as determined by a panel of experts such as officials' coaches (Larkin et al., 2014a; Spitz et al., 2018). Rugby union referees have been compared to touch-judges, assessors, and referee coaches (Mascarenhas et al., 2005a). The construct validity of this study was quite high, with the highest ranked referees scoring the best on the test. Interestingly though, referee coaches who are considered subject matter experts scored the lowest. Referees, coaches and players have been found to award more fouls when there is crowd noise (Balmer et al., 2007). These results, however, were presented with no differentiation between groups or accuracy of the decisions. There are a multitude of studies (Krenn, 2014; Renden et al., 2014; Spitz et al., 2017) comparing officials' decision-making to other populations based on a number of factors [such as uniform colour (Krenn, 2014), motor experience of playing the sport (Renden et al., 2014), and slow motion footage (Spitz et al., 2017)]. These studies only further demonstrate skill-based differences associated with the expertise effect, and video-based activities could be used to promote recruitment of these individuals into the ranks of officials. Upon analysis across this research area, there is a wide discrepancy of methods applied, especially in relation to decision-making accuracy. For results to have significance, reporting of decision-making accuracy within the video-based task is a necessity.

In terms of clip type used, the majority of studies (n = 23) have utilised broadcast footage as the presentation method. This is typically filmed from a fixed location in the grandstand and does not replicate the perspective of an official in a game (Craig, 2013), and as such lacks fidelity. Fidelity refers to the extent a situation replicates reality, and is a key element of transfer of on-field performance to off-field such as in video-based tasks (Alessi, 1988; Farrow, 2013; Lorains, Ball, & MacMahon, 2013). Wagner-Egger et al. (2012) presented decision-making scenarios from the video game FIFA 2005, with the rationale that this allows controlled scenarios from multiple viewpoints. As this task presents animations, however, it potentially lacks fidelity as it does not provide real world footage. To increase fidelity through heightened representativeness of the task, one study filmed match clips from the referees' perspective using head-mounted glasses (Nazarudin et al., 2015). Similarly, research has filmed match simulations from the perspective of assistant referees (Spitz et al., 2016; Spitz et al., 2018). These three studies have provided an advancement of the literature, by implementing first-

person/egocentric viewpoints to increase representativeness (Petit & Ripoll, 2008). A limitation which remains in this method is video footage does not automatically update with visual changes from head movements, which limited perception-action coupling (Craig, 2013). Virtual reality presentation of 360° videos is a possible technology to overcome this barrier to increase representativeness of the task, as it allows an individual's head movements to be the visual changes perceived. It is imperative for the video-based task to be representative of the match environment to determine differences between expertise levels (Williams & Ericsson, 2005). For this research area to advance, different technologies such as virtual reality should be compared to traditional methods such as broadcast footage.

Several studies have attempted to increase the representativeness of video-based tasks by introducing potential stressors or examining certain biases that may influence an official in a game. As Australian football umpires have a physically demanding task (Elsworthy et al., 2014), studies examined the effect of physical demands in relation to decision-making performance. These studies reported no relationship between physical exertion and video-based decision-making performance (Larkin et al., 2014b; Paradis et al., 2016), similar to match data (Elsworthy et al., 2014). As sport officials, especially elite, contend with significant crowd noise, this has been examined in the literature. One study involved soccer referees watching a game with fans or no fans present, with results demonstrating participants are more likely to disagree with the decision when there are no supporters present (Nevill et al., 2017). Introduction of match noise (i.e., players, commentary, and crowd) in the task leads to a home team bias in decision-making in two studies (Balmer et al., 2007; Nevill et al., 2002), but not necessarily another study (Lex et al., 2015). Despite the lack of influence match noise had on awarding fouls, Lex et al. (2015) reported that it led to a bias in awarding penalties (i.e., red or yellow card) against the away side. This supports the literature stating crowd noise does influence on-field decision-making (Downward & Jones, 2007; Goumas, 2014). Unlike crowd noise, there appears to be no decision-making bias against a team with an aggressive reputation (Jones et al., 2002). The influence of previous decisions has been examined, with results inferring this does indeed impact subsequent decisions in soccer (Plessner & Betsch, 2001), and basketball officials (Brand et al., 2006). Extending on this research, Poolton et al. (2011) reported referees who ruminate over previous decisions

are significantly more likely to award more fouls against the away team, supporting the home advantage theory also. The decision-making of referees in comparison to players and fans has been examined in relation to the skin colour of the soccer player. Wagner-Egger et al. (2012) highlighted there were no decision-making differences between these groups, but challenges made by black players were more likely to be considered fouls, whereas fouls made by white players were considered to be more severe. Uniform colour has also been examined as a possible decision-making bias as per skin colour. Results suggested soccer referees judged tackles less harshly than university students, but there was no overall impact of colour on tackle judgement (Krenn, 2014). Unfortunately though, the results of these two do not present the accuracy of the decisions. Although these findings indicate players of a specific shirt colour for example may commit "harsher" fouls, it is imperative to investigate whether these decisions are more or less accurate. The results suggest decision-making in off-field tasks can be influenced in the same way as on-field performance is, and certain biases can influence decision-making.

To accurately assess the performance of an individual on a video-based test, it is important to measure the transfer to match performance. In this review, however, only one study included the assessment of transfer, by calculating the correlation between video-based test performance and match performance (Nazarudin et al. 2015). Research has suggested including this integral component in video-based tasks for sports officials (Paradis et al., 2016), however this is rarely examined. In athletes, transfer tasks have been used to assess the effectiveness of perceptual-cognitive interventions (Lorains et al., 2013). Results suggest the introduction of a perceptual-cognitive training stimulus may not directly benefit on-field performance, despite performance improvement in an off-field video-based test (Gorman & Farrow, 2009; Lorains et al., 2013), suggesting a low correlation between on-field performance, and off-field video-based tests. Transfer tests have typically not been used in the research due to the inherent difficulties of natural in-game variation, and the lack of control researchers have over this environment. As a result, researchers typically do not incorporate this component as it leads to validity and reliability issues (Larkin et al., 2015). Despite this, a novel transfer test is better than none as sporting officials, and athletes in general, are ultimately measured by on-field performance. Only one study included in this review examined concurrent validity (Nazarudin et al. 2015).

Concurrent validity can be assessed via skill transfer by correlating the video-based test decision-making accuracy to on-field decision-making accuracy. Alternatively, performance ranking of officials could be assessed, as decision-making is the most important skill for officiating, and a key cornerstone of success (Helsen & Bultynck, 2004; Mascarenhas et al., 2005b). To advance this area of the literature, comparisons to on-field performance are vital.

3.5.1. Conclusions

In summary, this review has highlighted several key findings among the literature assessing decision-making skill in interactor officials. Firstly, soccer referees are the most predominantly investigated group of sporting officials. There were 16 studies which researched this group, compared to the next highest being Australian football and rugby union (three studies each). Video-based methods appear to have high construct validity when contrasting decision-making skill of multiple skill levels. Therefore, video-based tests can be used as a consistent and accurate measure of individual decision-making progression over a period of time. There is, however, a high amount of variability (i.e., clip type, number of clips, participating groups) in the methods across the studies identified which may lead to a certain degree of incompatibility among the findings. Finally, this review highlighted the influences which can bias officials' decision-making. Of these influences, crowd noise appears to be a prominent influence on decision-making accuracy.

3.5.2. Consideration for future research

Based on the key findings outlined above, the authors have several considerations for the direction of research in this area. As discussed, there is a plethora of research investigating decision-making of soccer referees. While it is recognised that is considered the world's most popular sport (Giulianotti, 2012), there is a necessity for research in other interactor officials to promote effective practice in other sports. In addition, validity and, especially, reliability measures need to be implemented in studies of this nature to ensure rigor of the video-based assessment tool. Similar to

consistency in the applied methodology of each study, this will ensure compatibility of results across different sports. The most common modality of video presentation was match broadcast footage of sporting games. With the advent of technology, other modalities such as virtual or augmented reality could be considered to be more representative of the in-game decision-making of a sporting official in an off-field controlled environment. For this research area to develop, the assessment of transfer is imperative. Although there are inherent limitations in examining transfer, the exploration of novel transfer assessments will further advance this area of the literature, hence reinforcing the practical implications of this method.

3.6. References

Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-based Instruction, 15*, 40-47.

Araújo, D., Davids, K., & Hristovski, R. (2006). The ecological dynamics of decision making in sport. *Psychology of Sport and Exercise, 7*(6), 653-676. https://doi.org/10.1016/j.psychsport.2006.07.002

Baker, J., Cote, J., & Abernethy, B. (2003). Sport-specific practice and the development of expert decision-making in team ball sports. *Journal of Applied Sport Psychology, 15*(1), 12-25. https://doi.org/10.1080/10413200305400

Baker, J., Côté, J., & Abernethy, B. (2003). Learning from the experts: Practice activities of expert decision makers in sport. *Research Quarterly for Exercise and Sport, 74*(3), 342-347. https://doi.org/10.1080/02701367.2003.10609101

Balmer, N. J., Nevill, A. M., Lane, A. M., & Ward, P. (2007). Influence of crowd noise on soccer refereeing consistency in soccer. *Journal of Sport Behavior, 30*(2), 130.

Berry, J., Abernethy, B., & Côté, J. (2008). The contribution of structured activity and deliberate play to the development of expert perceptual and decision-making skill. *Journal of Sport and 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*(1), 93-99.

Burnett, A. M., Bishop, D. T., Ashford, K. J., Williams, A. M., & Kinrade, N. P. (2017). Decision-making of English Netball Superleague umpires: Contextual and dispositional influences. *Psychology of Sport and Exercise, 31*, 52-60. http://dx.doi.org/10.1016/j.psychsport.2017.04.003

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

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

Craig, C. (2013). Understanding perception and action in sport: how can virtual reality technology help? *Sports Technology, 6*(4), 161-169. https://doi.org/10.1080/19346182.2013.855224

Cunningham, I., Simmons, P., Mascarenhas, D., & Redhead, S. (2014). Skilled interaction: concepts of communication and player management in the development of sport officials. *International Journal of Sport Communication, 7*(2), 166-187. https://doi.org/10.1123/IJSC.2013-0098

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26): Sage publications.

Downward, P., & Jones, M. (2007). Effects of crowd size on referee decisions: Analysis of the FA Cup. *Journal of Sports Sciences, 25*(14), 1541-1545. https://doi.org/10.1080/02640410701275193

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

Elsworthy, N., & Dascombe, B. J. (2011). The match demands of Australian rules football umpires in a state-based competition. *International Journal of Sports Physiology and Performance, 6*(4), 559-571. https://doi.org/10.1123/ijspp.6.4.559

Emmonds, S., O'Hara, J., Till, K., Jones, B., Brightmore, A., & Cooke, C. (2015). Physiological and Movement Demands of Rugby League Referees: Influence on Penalty Accuracy. *Journal of Strength and Conditioning Research, 29*(12), 3367-3374. https://doi.org/10.1519/JSC.0000000000001002

Farrow, D. (2013). Practice-enhancing technology: a review of perceptual training applications in sport. Gadotti, I., Vieira, E., & Magee, D. (2006). Importance and clarification of measurement properties in rehabilitation. *Brazilian Journal of Physical Therapy, 10*(2), 137-146. https://doi.org/10.1590/S1413-35552006000200002

Ghasemi, A., Momeni, M., Jafarzadehpur, E., Rezaee, M., & Taheri, H. (2011). Visual skills involved in decision making by expert referees. *Perceptual and Motor Skills, 112*(1), 161-171. https://doi.org/10.2466%2F05.22.24.27.PMS.112.1.161-171

Giulianotti, R. (2012). *Football*: Wiley Online Library.

Gorman, A. D., & Farrow, D. (2009). Perceptual training using explicit and implicit instructional techniques: Does it benefit skilled performers? *International Journal of Sports Science and Coaching, 4*(2), 193-208. https://doi.org/10.1260%2F174795409788549526

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

Hancock, D. J., Rix-Lièvre, G., & Côté, J. (2015). Citation network analysis of research on sport officials: A lack of interconnectivity. *International Review of Sport and Exercise Psychology, 8*(1), 95-105. https://doi.org/10.1080/1750984X.2015.1022202

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247. https://psycnet.apa.org/doi/10.1037/1040-3590.7.3.238

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

Johansen, B. T., & 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.

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

Krenn, B. (2014). The impact of uniform color on judging tackles in association football. *Psychology of Sport and Exercise, 15*(2), 222-225. https://doi.org/10.1016/j.psychsport.2013.11.007

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

Larkin, P., Berry, J., Dawson, B., & Lay, B. (2011). Perceptual and decision-making skills of Australian football umpires. *International Journal of Performance Analysis in Sport, 11*(3), 427-437. https://doi.org/10.1080/24748668.2011.11868562

Larkin, P., Mesagno, C., Berry, J., & Spittle, M. (2014a). Development of a valid and reliable video-based decision-making test for Australian football umpires. *Journal of Science and Medicine in Sport, 17*(5), 552-555. https://doi.org/10.1016/j.jsams.2013.08.001

Larkin, P., Mesagno, C., Berry, J., & Spittle, M. (2018a). Exploration of the perceptual-cognitive processes that contribute to in-game decision-making of Australian football umpires. *International Journal of Sport and Exercise Psychology, 16*(2), 112-124. https://doi.org/10.1080/1612197X.2016.1167760

Larkin, P., Mesagno, C., Spittle, M., & Berry, J. (2015). An evaluation of video-based training programs for perceptual-cognitive skill development. A systematic review of current sport-based knowledge. *International Journal of Sport Psychology, 46*(6), 555-586.

Larkin, P., O'Brien, B., Mesagno, C., Berry, J., Harvey, J., & Spittle, M. (2014b). Assessment of decision-making performance and in-game physical exertion of Australian football umpires. *Journal of Sports Sciences, 32*(15), 1446-1453. https://doi.org/10.1080/02640414.2014.896998

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

Lorains, M., Ball, K., & MacMahon, C. (2013a). An above real time training intervention for sport decision making. *Psychology of Sport and Exercise, 14*(5), 670-674. https://doi.org/10.1016/j.psychsport.2013.05.005

MacMahon, C., Helsen, W., Starkes, J., & Weston, M. (2007a). Decision-making skills and deliberate practice in elite association football referees. *Journal of Sports Sciences, 25*(1), 65-78. https://doi.org/10.1080/02640410600718640

MacMahon, C., Mascarenhas, D., Plessner, H., Pizzera, A., Oudejans, R., & Raab, M. (2014). *Sports officials and officiating: Science and practice*. Oxon: Routledge.

MacMahon, C., Starkes, J., & Deakin, J. (2007b). Referee decision making in a video-based infraction detection task: Application and training considerations. *International Journal of Sports Science and Coaching, 2*(3), 257-265. https://doi.org/10.1260%2F174795407782233164

MacMahon, C., & Ste-Marie, D. (2002). Decision-making by experienced rugby referees: Use of perceptual information and episodic memory. *Perceptual and Motor Skills, 95*(2), 570-572. https://doi.org/10.2466%2Fpms.2002.95.2.570

Mallo, J., Frutos, P. G., 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

Marteniuk, R. G. (1976). *Information processing in motor skills*. New York: Holt, Rinehart and Winston.

Mascarenhas, D., Collins, D., & Mortimer, P. (2005a). The accuracy, agreement and coherence of decision-making in rugby union officials. *Journal of Sport Behavior, 28*(3), 253-271.

Mascarenhas, D., Collins, D., & Mortimer, P. (2005b). Elite Refereeing Performance: Developing a Model for Sport Science Support. *Sport Psychologist, 19*, 364-379. https://doi.org/10.1123/tsp.19.4.364

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Morris, G., & O'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

Nazarudin, M. N., Abdullah, M. R., Suppiah, P. K., Fauzee, M. S. O., Parnabas, V., & Abdullah, N. M. (2015). Decision making and performance of Malaysian rugby sevens referees. *Movement, Health and Exercise, 4*(1), 61-75. https://doi.org/10.15282/mohe.v4i0.45

Nevill, A. M., Balmer, N. J., & Williams, A. M. (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

Nevill, A. M., Hemingway, A., Greaves, R., Dallaway, A., & Devonport, T. J. (2017). Inconsistency of decision-making, the Achilles heel of referees. *Journal of Sports Sciences, 35*(22), 2257-2261. https://doi.org/10.1080/02640414.2016.1265143

Neville, T. J., Salmon, P. M., & Read, G. J. (2018). Radio Gaga? Intra-team communication of Australian Rules Football Umpires–Effect of radio communication on content, structure and frequency. *Ergonomics, 61*(2), 313-328. https://doi.org/10.1080/00140139.2017.1353140

Page, K., & Page, L. (2010). Alone against the crowd: Individual differences in referees' ability to cope under pressure. *Journal of Economic Psychology, 31*(2), 192-199. https://doi.org/10.1016/j.joep.2009.08.007

Paradis, K., Larkin, P., & O'Connor, D. (2016). The effects of physical exertion on decision-making performance of Australian football umpires. *Journal of Sports Sciences, 34*(16), 1535-1541. https://doi.org/10.1080/02640414.2015.1122205

Petit, J.-P., & Ripoll, H. (2008). Scene perception and decision making in sport simulation: a masked priming investigation. *International Journal of Sport Psychology, 39*(1), 1-19.

Plessner, H., & Betsch, T. (2001). Sequential effects in important referee decisions: The case of penalties in soccer. *Journal of Sport and Exercise Psychology, 23*(3), 254-259. https://doi.org/10.1123/jsep.23.3.254

Poolton, J., Siu, C. M., & Masters, R. (2011). The home team advantage gives football referees something to ruminate about. *International Journal of Sports Science and Coaching, 6*(4), 545-552. https://doi.org/10.1260%2F1747-9541.6.4.545

Renden, P. G., Kerstens, S., Oudejans, R. R., & Cañal-Bruland, R. (2014). Foul or dive? Motor contributions to judging ambiguous foul situations in football. *European Journal of Sport Science, 14*(sup1), S221-S227. https://doi.org/10.1080/17461391.2012.683813

Souchon, N., Coulomb-Cabagno, G., Traclet, A., & Rascle, O. (2004). Referees' decision making in handball and transgressive behaviors: Influence of stereotypes about gender of players? *Sex Roles, 51*(7-8), 445-453. https://doi.org/10.1023/B:SERS.0000049233.28353.f0

Souchon, N., Livingstone, A. G., & Maio, G. R. (2013). The influence of referees' expertise, gender, motivation, and time constraints on decisional bias against women. *Journal of Sport and Exercise Psychology, 35*(6), 585-599. https://doi.org/10.1123/jsep.35.6.585

Spitz, J., Put, K., Wagemans, J., Williams, A. M., & Helsen, W. F. (2016). Visual search behaviors of association football referees during assessment of foul play situations. *Cognitive Research: Principles and Implications, 1*(1), 1-11. https://doi.org/10.1186/s41235-016-0013-8

Spitz, J., Put, K., Wagemans, J., Williams, A. M., & Helsen, W. F. (2017). Does slow motion impact on the perception of foul play in football? *European Journal of Sport Science, 17*(6), 748-756. https://doi.org/10.1080/17461391.2017.1304580

Spitz, J., Put, K., Wagemans, J., Williams, A. M., & Helsen, W. F. (2018). The role of domain-generic and domain-specific perceptual-cognitive skills in association football referees. *Psychology of Sport and Exercise, 34*(Supplement C), 47-56. https://doi.org/10.1016/j.psychsport.2017.09.010

Swann, C., Moran, A., & Piggott, D. (2015). Defining elite athletes: Issues in the study of expert performance in sport psychology. *Psychology of Sport and Exercise, 16, Part 1*, 3-14. http://dx.doi.org/10.1016/j.psychsport.2014.07.004

Thomas, J. R., Silverman, S., & Nelson, J. (2015). *Research methods in physical activity, 7E*: Human Kinetics.

Wagner-Egger, P., Gygax, P., & Ribordy, F. (2012). Racism in soccer? Perception of challenges of black and white players by white referees, soccer players, and fans. *Perceptual and Motor Skills, 114*(1), 275-289. https://doi.org/10.2466%2F05.07.17.PMS.114.1.275-289

Williams, A. 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

Williams, A. M., Ward, P., Smeeton, N., & Allen, D. (2004). Developing anticipation skills in tennis using on-court instruction: Perception versus perception and action. *Journal of Applied Sport Psychology, 16*(4), 350-360. https://doi.org/10.1080/10413200490518002

Wilson, A. W., & Mock, S. E. (2013). Association of certification level and assertiveness with accuracy of calls among ice hockey referees. *International Journal of Sports Science and Coaching, 8*(3), 505-512. https://doi.org/10.1260%2F1747-9541.8.3.505

Chapter 4: Reliability & Validity of 360° VR and Match Broadcast Video

This Chapter is presented in pre-publication format of a recent publication titled:

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. (2019c). Using 360° virtual reality as a decision-making assessment tool in sport. *Journal of Science and Medicine in Sport*, 22(9), 1049-1053. https://doi.org/10.1016/j.jsams.2019.03.012

VICTORIA UNIVERSITY

THE NEW WAY TO DO UNI

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book:	**Kittel, A.**, Larkin, P., Elsworthy, N., & Spittle, M. (2019c). Using 360° virtual reality as a decision-making assessment tool in sport. *Journal of Science and Medicine in Sport*, 22(9), 1049-1053. https://doi.org/10.1016/j.jsams.2019.03.012

Surname:	Kittel	First name:	Aden
Institute:	Institute for Health and Sport	Candidate's Contribution (%):	80

Status:

Accepted and in press:	Date:	
Published:	Date:	Sep 2019

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be included in the thesis as outlined in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted]	12/8/2020
Signature	**Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

PO Box 14428, Melbourne,
Vic 8001, Australia
+61 3 9919 6100

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location(s)**:

All electronic data will be stored on the Victoria University R Drive, which is a secure research storage drive.

Name(s) of Co-Author(s)	Contribution (%)	Nature of Contribution	Signature	Date
Paul Larkin	8	Reviewer, provided feedback, minor edits.	[redacted]	12/8/20
Nathan Elsworthy	6	Reviewer, provided feedback, minor edits.		12/8/20
Michael Spittle	6	Reviewer, provided feedback, minor edits.		12/8/20

Updated: September 2019

4.1. Abstract

Objectives: To examine the reliability, construct validity and ecological validity of 360° VR and match broadcast footage for off-field decision-making assessment in Australian football umpires. *Design*: Validation assessments with test re-test reliability. *Methods*: Two video-based tests of 60 clips each were developed to assess Australian football umpire decision-making, including 360° video of small-sided Australian football games and match broadcast footage of AFL games. Elite (n = 13) and amateur (n = 15) umpires participated in two testing sessions, in a randomised, counterbalanced design. Test re-test reliability was assessed using Cohen's Kappa for individual clips and Intraclass Correlation Coefficients for test scores. Video tests were assessed for construct validity. Ecological validity of the decision-making processes was assessed for each method. *Results*: 31 clips met the minimum Kappa criteria for the 360° VR test and 28 clips for match broadcast. Results indicated strong reliability for the 360° VR (ICC = 0.89) and match broadcast (ICC = 0.89) tests. For both video modes, elite umpires performed significantly better in decision-making accuracy than amateur ($p < 0.05$). For ecological validity of the decision-making processes, 360° VR was rated significantly higher than match broadcast vision ($p < 0.05$) overall. *Conclusions*: This is the first study to examine the reliability and validity of 360° VR footage as an off-field decision-making assessment tool in sport. As match broadcast vision is commonly used to assess decision-making in athletes and officials, results suggest that 360° VR is also an appropriate assessment tool. Although both video modes demonstrate similar reliability and construct validity, 360° VR was considered more specific to in-game decision-making processes, suggesting stronger ecological validity.

Keywords: Perceptual-cognitive skill, virtual reality, decision-making assessment, test validation.

4.2. Introduction

Decision-making is a key perceptual-cognitive skill in sport, defined as the ability to perceive and correctly interpret game-related information to select an appropriate sport-specific response (Baker, Côté, & Abernethy, 2003). There has been extensive research investigating decision-making processes of athletes (Araújo, Davids, & Hristovski, 2006; Baker et al., 2003). Importantly, decision-making skill can differentiate between skilled and less-skilled performers in sport (Berry, Abernethy, & Côté, 2008; Williams & Ericsson, 2005). Sporting officials have emerged as a key group to examine this skill as decision-making is the most important skill for successful performance (Helsen & Bultynck, 2004; Larkin, Berry, Dawson, & Lay, 2011). It has been acknowledged that to develop decision-making, the best form of practice is in-game performance (MacMahon, Helsen, Starkes, & Weston, 2007). There are only a finite amount of games with inherent physical loads, which has led to the development of additional methods to investigate decision-making. To investigate and assess decision-making, video-based methods are commonly used as they provide a controlled off-field environment for testing and training both athletes (Lorains, Ball, & MacMahon, 2013) and officials (Larkin, Mesagno, Berry, & Spittle, 2014). Due to the high variability of match decision-making, video-based assessments provide a reliable and valid measure of decision-making performance.

An advantage of video-based methods is they provide greater control and consistency of scenarios than in-game decision-making assessments (Larkin et al., 2014). Video-based methods present a short game-based scenario where an athlete or official must make a sport-specific decision. Video-based measures have highlighted differences between skilled and less-skilled performers in a range of athletes such as Australian football players (Lorains et al., 2013) and Australian football umpires (Larkin et al., 2014; Paradis, Larkin, & O'Connor, 2016), suggesting strong validity of the video-based tasks. There are, however, inherent limitations associated with using video-based methods. For example, these methods typically use a broadcast or third-person (exocentric) perspective, which is filmed from an elevated position in the grandstand (Craig, 2013). As a result, footage from an exocentric perspective does not replicate the position or the potential in-game perceptual-cognitive information required in the decision-making process of an umpire or athlete. Research has suggested the

implementation of first-person (egocentric) viewpoints would increase the representativeness of decision-making tasks (Petit & Ripoll, 2008). Further, traditional video-based methods also limit interactivity and representativeness of the task whereby video footage does not automatically update with visual changes from head movements (Craig, 2013). This is critical in 360° sports such as soccer, basketball, and Australian football as it could provide information regarding the ability to identify player movement patterns in all areas of the field. Despite a range of research using video-based methods, there is potential for researchers to create more representative tasks.

To address these limitations, virtual reality has been suggested as a more suitable and representative task to assess decision-making (Craig, 2013; Farrow, Reid, Buszard, & Kovalchik, 2018). Virtual reality is a term used to describe simulations of a real or imaginary environment, whereby a participant can both perceive and move (interact) within the environment (Craig, 2013). Typically, virtual reality has explored animated/virtual sporting environments presented on a large screen (Vignais, Kulpa, Brault, Presse, & Bideau, 2015), or through a head mounted display (HMD) (Brault, Bideau, Kulpa, & Craig, 2012). A key premise of using virtual reality is to increase representativeness, suggesting real world footage may be a suitable alternative to animated footage when using this technology. 360° VR has emerged as a possible testing and training tool for decision-making in sport. A recent study examined the effectiveness of 360° VR on enhancing decision-making in basketball players (Panchuk, Klusemann, & Hadlow, 2018). This study found large yet non-significant improvements following the intervention, suggesting more research is required to determine the effectiveness of this technology. The reliability, construct validity and ecological validity of 360° video viewed through a HMD as a decision-making assessment tool, however, has not yet been examined in the literature. Therefore, the primary of aim of this study is to examine both the reliability and construct validity of two different video-based methods to assess Australian football umpires' decision-making skill. These two video-based methods are 360° VR (i.e., egocentric viewpoint) and traditional match broadcast footage (i.e., exocentric viewpoint). A secondary aim will be to evaluate the level of ecological validity of each method through game-likeness ratings.

4.3. Methods

28 participants across two performance levels, elite and amateur, participated in this study. The elite group was comprised of 13 Australian football umpires who officiated in the Australian Football League (AFL), with a mean age of 29.9 years (SD = 5.9). The elite cohort had umpired for an average of 15.1 years, with an average of 7.0 years at the AFL competition level. At the time of this study, the average AFL matches officiated by this group was 118 (SD = 109), with all umpires being on the elite list for at least one full season. The amateur group consisted of 15 Australian football umpires who officiated in an Australian football Metropolitan division one competition, with an average age of 27.8 years (SD = 12.0). These participants had umpired Australian football for an average of 9.7 years, with none of this group having umpired at AFL level. Ethics approval was granted by the Victoria University Human Research Ethics Committee (approval number: HRE18-122) and all participants provided informed consent prior to participating in the study.

Two different video-based methods for decision-making skill of Australian football umpires were developed: 360° VR footage and match broadcast footage. For both tests, inclusion criteria for decisions were player contact scenarios, with decisions including: “play on”, “holding the ball”, “high tackle”, “push in the back” and “contact below the knees” as outlined by the “Decision-making heuristic for a tackling situation” in Australian football (Larkin, Mesagno, Berry, & Spittle, 2018a).

360° VR footage was obtained from small-sided games and match simulations completed by high performance Under 18 Australian football players as part of their standard training procedures. This was filmed from the perspective of an umpire in match play using a 360° video camera (360fly 4k Camera, 360fly, Canonsburg, USA). The camera was positioned on a tripod 1.5m above the ground. To optimise representativeness of the footage, the camera was positioned approximately 15m from the training drills, in a similar position to an umpire in-game. Figure 4.1 presents a static example of this footage and a map of the setup.

For the match broadcast test, footage from the 2017 AFL Premiership Season was obtained from the AFL Umpiring Department. This footage was filmed from an elevated position in the

grandstand. Videos were not included if the elite cohort had received group coaching on the particular scenario, or if the umpire's in-game decision was observable in the clip. Figure 4.1 presents a static example of this footage and a map of the setup.

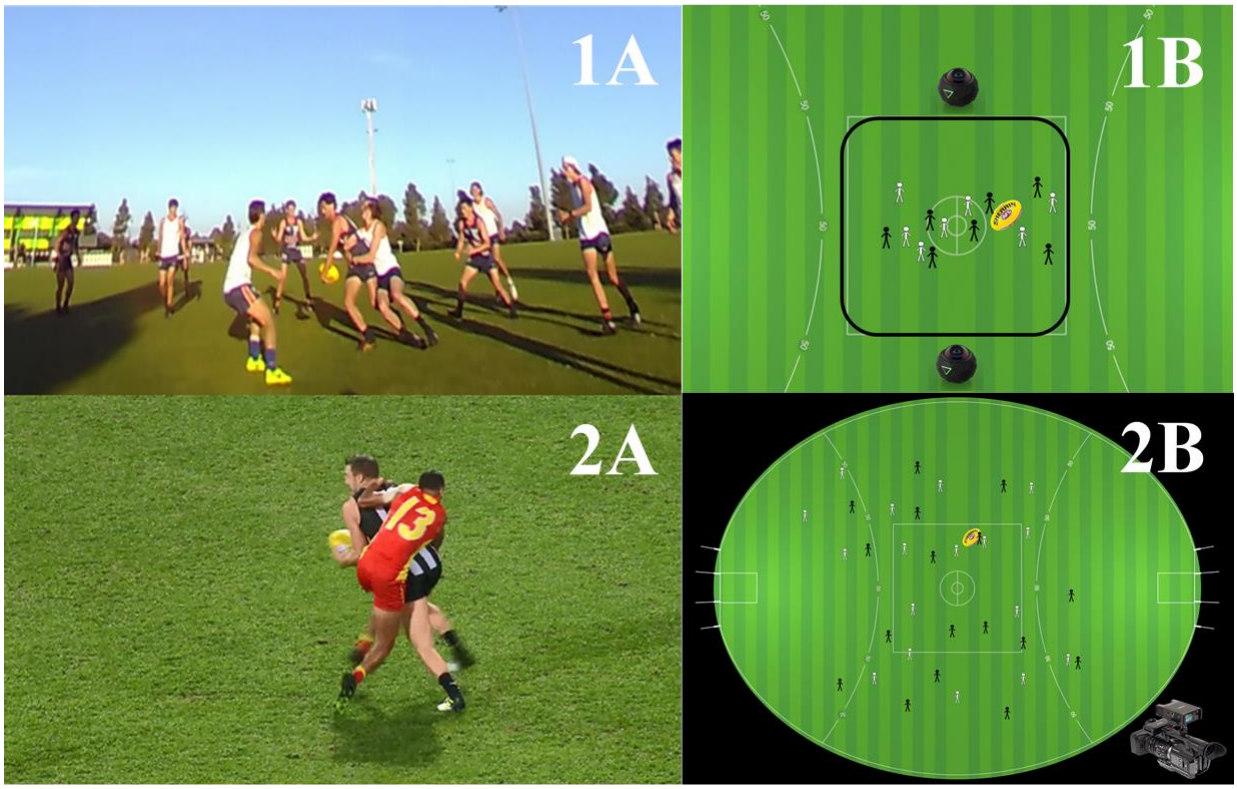

Figure 4.1: An example of the 360° VR and match broadcast footage and design.

(1A) 360° VR decision-making test footage, (1B) 360° VR filming design, (2A) match broadcast decision-making test footage and (2B) match broadcast filming design.

For each mode, clips were edited into short (3 - 12 seconds) individual clips presenting a possible infringement (i.e., free kick) as per previous methods implementing a video-based decision-making task (Kittel, Elsworthy, & Spittle, 2019). Clips were excluded if there was a poor/distant camera angle. 100 clips from each video mode were presented to subject matter experts: full-time AFL umpire coaches (n = 2). The two coaches had on average 10 years' experience coaching at the elite AFL level. The coaches assessed the face validity (i.e., whether there was a potential decision-making scenario) and the correct response for each clip. If the coaches did not agree, a discussion on the correct decision occurred until they came to a consensus decision. At this stage, videos were removed if there was no

clear decision, per coaches' feedback. The coaches and the authors discussed the final selection of clips to be used in both decision-making tests. To ensure consistency between tests, there was an identical number of each individual decision type (i.e., "play on", "holding the ball", "high tackle", "push in the back" and "contact below the knees") within each test. "Play on" decisions, which were situations where no free kicks were awarded but still required a deliberate "no" decision from the umpire (Kittel et al., 2019) were also included within the tests.

For both tests, a three second period was presented before each clip indicating the clip number in white text on a black screen. For the 360° VR condition, this period also included arrows pointing to where the clip number was, to enable the participants to orientate themselves within the 360° space. This was followed by the presentation of a static image of the first frame of the clip for two seconds to provide context, immediately followed by the clip. At the conclusion of each clip, a static image of the final frame appeared for 0.5 seconds. This was followed by a six second period where a black screen was presented, with white text stating "make your decision" per previous methods (Larkin et al., 2014). This sequence was repeated for all clips. Audio was removed from all clips in both conditions. Final clips were rendered in 4k video (3840 x 1920 pixels) for 360° VR condition and Full HD (1920 x 1080 pixels) for match broadcast video using ADOBE Premiere Pro CS4 video editing package.

Participants completed the decision-making tests two weeks apart at the beginning of a scheduled training session, prior to any physical activity. Standardised conditions were followed for both testing sessions. Each testing session included the 360° VR decision-making task (60 clips) and the match broadcast decision-making task (60 clips), with the sequence of test administration randomised for each participant. Both modes included three practice clips prior to the test to familiarise the participants with the process. These clips were not repeated in the test battery. The 360° VR decision-making task was presented on an Oculus Go HMD (Oculus Go, Oculus, California, USA). For the 360° VR task, participants were seated in a swivel chair to allow the participants to freely and safely view the 360° vision. Participants were required to verbalise their decision for the 360° VR task as they would in a game, with their responses recorded by the research team. The match broadcast test was presented on iPads in a quiet room (Apple iPad Air 2, Apple Inc., Cupertino, USA) as per previous

methods (Kittel et al., 2019). Participants provided written responses on an answer sheet (i.e., “play on”, “holding the ball”, “high tackle”, “push in the back” and “contact below the knees”) (Larkin et al., 2014). For both tests, clip order was randomised between participants and testing sessions.

4.3.1. Reliability analysis

Test re-test reliability of the individual video-based clip responses was examined by Cohen’s Kappa, as per previous research (Larkin et al., 2014). Kappa values were categorised as poor agreement (< 0.00), slight agreement (0.00 – 0.20), fair agreement (0.21 – 0.40), moderate agreement (0.41 – 0.60), substantial agreement (0.61 – 0.80), almost perfect agreement (0.81 – 1.00) (Landis & Koch, 1977). Values greater than 0.40 (moderate agreement) were deemed reliable, providing a stronger assessment of reliability than previous research using 0.30 as the benchmark (Larkin et al., 2014). Of the clips deemed reliable (i.e., kappa values > 0.40), Intraclass Correlation Coefficients (ICC) were calculated for 360° VR and match broadcast video modes, by comparing the scores of each test across the two test sessions (i.e., comparing 360° VR test 1 to 360° VR test 2) (Spitz, Put, Wagemans, Williams, & Helsen, 2017). This was calculated through a through a widely available spreadsheet (Hopkins, 2015). ICC values were classified as the following: poor reliability (< 0.50), moderate reliability (0.50 – 0.74), good reliability (0.75 – 0.89) and excellent reliability (≥ 0.90) (Koo & Li, 2016).

4.3.2. Construct validity analysis

Decision-making accuracy was calculated by the number of clips with the correct decision response, as determined by subject matter experts. Decision-making differences were calculated only for the clips deemed reliable (kappa value > 0.40). Decision-making accuracy differences between the two groups (elite vs. amateur) were assessed for each video mode, which is a form of construct validity by differentiating between known skill levels (Gadotti, Vieira, & Magee, 2006). To assess decision-making differences between the two groups an independent samples t-test was performed for each video mode, with an alpha level of 0.05 deemed significant. Each video method was assessed for normality

using the Shapiro-Wilks test. Effect sizes were calculated using Cohen's *d* statistic. Benchmarks for the effect sizes (*d*) were small (0.20 - 0.49), moderate (0.50 - 0.79) and large (≥ 0.80) (Cohen, 1992).

4.3.3. Ecological validity analysis

The ecological validity (Araujo, Davids, & Passos, 2007) of each video mode was determined by how closely each video mode replicated the decision-making processes of an umpire in-game. Following the second testing session, participants completed a visual analogue scale (Catteeuw, Gilis, Jaspers, Wagemans, & Helsen, 2010) for each video mode, determining how much the participants perceived each video mode (i.e., 360° VR and match broadcast) to be like in-game decision-making. Participants were asked to "Make a dash on the line for how each method feels like your match decision-making process" using a 10cm visual analogue scale ranging from "nothing like match decision-making", to "exactly like match decision-making". Differences in game-likeness ratings between the video modes were assessed using paired sample t-tests. This was analysed by skill level (i.e., elite and amateur), and for the participants overall.

4.4. Results

4.4.1. Reliability

The set of 60 clips for each test were each individually analysed for test re-test reliability. For the 360° VR condition, results indicated that 31 clips were considered reliable (i.e., kappa values > 0.40). There were 28 clips meeting the minimum reliability criteria (i.e., kappa values > 0.40) in the match broadcast test. The ICC demonstrated good reliability for both the 360° VR test (ICC = 0.89) and the match broadcast decision-making test (ICC = 0.89), indicating consistent decision-making accuracy scores across the testing sessions.

4.4.2. Construct validity

The clips identified as reliable for both tests were assessed for construct validity. The average decision-making accuracy score (%) between the two testing sessions for each video mode was assessed for each participant (see Table 4.1). The elite umpires scored significantly higher than the amateur cohort in both the 360° VR test, $t(25) = 3.78$, $p < 0.01$ and the match broadcast test, $t(23) = 4.40$, $p < 0.01$. There were large effect sizes between groups for the 360° VR test ($d = 1.38$) and the match broadcast test ($d = 1.65$). Therefore, these results suggest strong construct validity of each assessment mode.

Table 4.1: Mean percentage score for the 360° VR and match broadcast decision-making tests.

	Mean (SD)		**F**	***p***	**Partial η^2**
	Elite	**Amateur**			
360° VR (%)					
	75.18 (6.37)	65.21 (7.29)	1.13	0.001	0.36
Match broadcast (%)					
	71.75 (10.65)	50.51 (12.90)	0.66	0.000	0.46

4.4.3. Ecological validity

The results from the game-likeness questionnaire indicated that overall (see Table 4.2), participants rated 360° VR (6.50 ± 1.65) as significantly more game-like than match broadcast footage (5.00 ± 2.52), $t(18) = 2.41$, $p = 0.03$, $d = 0.59$. When considering the game-likeness ratings for the elite umpires only, there were no significant differences between the 360° VR (5.59 ± 1.26) and match broadcast footage (4.63 ± 2.55), $t(10) = 1.084$, $p = 0.30$, $d = 0.33$. The amateur umpires considered 360° VR (7.75 ± 1.28) to be significantly more game like than match broadcast footage (5.50 ± 2.56), $t(7) = 2.679$, $p = 0.03$, $d = 0.95$.

Table 4.2: Ecological validity of the two video-based modes.

	Mean (SD)		*p*	Effect size
	360° VR	**Match broadcast**		
All participants			.03	.59
	6.50 (1.65)	5.00 (2.52)		
Elite only			.30	.33
	5.59 (1.26)	4.63 (2.55)		
Amateur only			.03	.95
	7.75 (1.28)	5.50 (2.56)		

4.5. Discussion

Numerous studies have suggested virtual reality to be a suitable tool to assess perceptual-cognitive skills such as decision-making (Craig, 2013; Farrow et al., 2018), with 360° VR suggested to be an appropriate training tool (Panchuk et al., 2018). This is the first study to investigate the reliability and validity of 360° VR footage to assess sport-specific decision-making skill. Results indicated 360° VR demonstrates strong reliability and validity, similar to previous research examining match broadcast footage with Australian football umpires (Larkin et al., 2014). Given the strong construct validity of each video mode, this infers 360° VR can be used as an alternative to traditional broadcast methods in assessing decision-making skill within sports officials.

In the literature, very few studies have examined test re-test reliability, in addition to the validity, of a video-based assessment tool (Larkin, Mesagno, Spittle, & Berry, 2015). It is integral that a test with strong validity must first have strong reliability, to ensure test differences are due to actual performance, rather than random test error (Gadotti et al., 2006). Similar research in Australian football umpires examined the test re-test reliability of a battery of clips, to be used in future testing scenarios within practice and research (Larkin et al., 2014). The strong reliability of the test overall, as evident by the ICC, demonstrate the reliability of using 360° VR as a decision-making assessment tool. Anecdotal results suggest that clips not deemed reliable were due to insufficient information presented in the clip, rather than the skill level of the participants. This study provides an updated battery of broadcast clips, in addition to 360° VR clips which are yet to be validated within the sporting literature.

This research contributes to the perceptual-cognitive literature investigating the off-field decision-making performance of sporting individuals. Construct validity is an imperative measure of video-based tasks, as it differentiates between known skill levels (Gadotti et al., 2006). When implementing a video-based test without this knowledge, it may be unclear whether differences are a result of the skill level differences, or the test design (Larkin et al., 2014). Elite officials scored significantly higher than amateur umpires in each video mode, inferring strong construct validity of both video modes. The results in the present study are analogous to previous video-based methods in Australian football umpires, where elite officials outperformed those from subsequent levels (Larkin et al., 2011). Therefore, results suggest 360° VR is an appropriate tool to elicit skill-based decision-making differences.

Virtual reality is broadly defined as a video simulation of a real or imaginary environment (Craig, 2013). Two key premises of using virtual reality are to increase representativeness of an off-field environment and secondly to increase immersion in the task through visual changes that coincide with head movements (Craig, 2013). The primary coaching method to improve decision-making skill within sports officials is to utilise video-based methods in an off-field environment. The findings of this study illustrate the application of 360° video to immerse the umpire in the game environment from a first-person perspective with positive results. Participants rated 360° VR to be more like their match decision-making processes as assessed through the game-likeness ratings, indicating stronger ecological validity of the decision-making processes of the task. This suggests 360° VR could be a more ecologically valid method to present in-game vision for decision-making assessment purposes. Using HMDs allows the participant to select where they are looking, at critical points throughout the clip. In previous first-person investigations not using 360° VR (Mann et al., 2009), the final footage is dependent upon the direction of the camera at the time of filming, therefore limiting the interactivity of the task. HMDs overcome this limitation and have been used to present virtual reality footage in predicting deceptive movement of rugby players using virtual footage (i.e., not footage of a 360° real world scenario) (Brault et al., 2012). Possibly due to the high interactivity of using a HMD, experts significantly outperformed novices in the task (Brault et al., 2012) suggesting this to be a more

representative presentation method, similar to 360° VR (Panchuk et al., 2018). Further, developing 360° VR scenarios compared to virtual computer environments is a more affordable process. When considering the accessibility and theoretical premises for using 360° VR along with the results of this study, it appears to be a suitable direction for sport-based decision-making research.

This study presents an opportunity for 360° VR to be examined further in the sporting literature. There is a significant research base utilising video-based methods to assess decision-making in sporting officials (Paradis et al., 2016) and players of different sports (Lorains et al., 2013; Plessner & Betsch, 2001). Although there have been promising findings from these studies using broadcast footage, the current study demonstrates 360° VR is a reliable and valid method to assess perceptual-cognitive performance. Given the proposal of 360° VR as an appropriate decision-making assessment tool, future researchers should also consider 360° VR as a training tool to develop decision-making performance of officials, in addition to athletes (Panchuk et al., 2018). There is a range of empirical studies examining the effectiveness of video-based training for the development of decision-making skill (Larkin et al., 2015), found to be an effective means of hastening expertise of sporting individuals, especially during the offseason or when an athlete is injured (Larkin, Mesagno, Berry, Spittle, & Harvey, 2018). As technologies such as 360° VR become more readily available, there is a need for researchers to further investigate how these methods may improve on-field performance. Future studies could utilise similar methods to this study to develop a selection of 360° VR clips for the purpose of assessing and training decision-making skill. Finally, the game-likeness rating used in this study only assessed the decision-making processes of the two technologies in relation to in-game decision-making performance. Future researchers should consider the impact of other factors such as the testing environment, response mode and decision-making scenarios (i.e., match footage vs. training situations) on the representativeness of the task. This knowledge would provide researchers and practitioners with a greater understanding of how to create a stronger representation of in-game decision-making using these technologies.

In conclusion, this is the first study to examine the use of 360° VR as a reliable and valid decision-making assessment tool in sport. The results highlight 360° VR is a reliable and valid assessment tool of decision-making in Australian football umpires, which is consistent with previously

used video-based broadcast methods. The higher game-likeness ratings of 360° VR compared to match broadcast footage suggests 360° VR is a stronger representation of in-game decision-making processes. Future investigations should explore the use of this modality in other sports and contexts.

4.5.1. Practical implications

- 360° VR is a reliable and valid assessment tool for perceptual-cognitive skills such as decision-making and has potential to be explored in a range of sporting environments.
- 360° VR is considered to represent more game-like decision-making processes than footage sourced from a broadcast perspective, suggesting 360° VR to be a more ecologically valid decision-making tool.
- Future studies should compare the effectiveness of both 360° VR and traditional broadcast methods in developing decision-making.

4.6. References

Araújo, D., Davids, K., & Hristovski, R. (2006). The ecological dynamics of decision making in sport. *Psychology of Sport and Exercise, 7*(6), 653-676. https://doi.org/10.1016/j.psychsport.2006.07.002

Araujo, D., Davids, K., & Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: Comment on Rogers, Kadar, and Costall (2005). *Ecological Psychology, 19*(1), 69-78. https://doi.org/10.1080/10407410709336951

Baker, J., Côté, J., & Abernethy, B. (2003). Learning from the experts: Practice activities of expert decision makers in sport. *Research Quarterly for Exercise and Sport, 74*(3), 342-347. https://doi.org/10.1080/02701367.2003.10609101

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

Brault, S., Bideau, B., Kulpa, R., & Craig, C. M. (2012). Detecting deception in movement: the case of the side-step in rugby. *PLoS ONE, 7*(6), e37494. https://doi.org/10.1371/journal.pone.0037494

Catteeuw, P., Gilis, B., Jaspers, A., Wagemans, J., & Helsen, W. (2010b). Training of perceptual-cognitive skills in offside decision making. *Journal of Sport and Exercise Psychology, 32*(6), 845-861. https://doi.org/10.1123/jsep.32.6.845

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98-101. https://doi.org/10.1111%2F1467-8721.ep10768783

Craig, C. (2013). Understanding perception and action in sport: how can virtual reality technology help? *Sports Technology, 6*(4), 161-169. https://doi.org/10.1080/19346182.2013.855224

Farrow, D., Reid, M., Buszard, T., & Kovalchik, S. (2018). Charting the development of sport expertise: challenges and opportunities. *International Review of Sport and Exercise Psychology, 11*(1), 238-257. https://doi.org/10.1080/1750984X.2017.1290817

Gadotti, I., Vieira, E., & Magee, D. (2006). Importance and clarification of measurement properties in rehabilitation. *Brazilian Journal of Physical Therapy, 10*(2), 137-146. https://doi.org/10.1590/S1413-35552006000200002

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

Hopkins, W. G. (2015). Spreadsheets for analysis of validity and reliability. *Sportscience, 19*, 36-44.

Kittel, A., Elsworthy, N., & Spittle, M. (2019a). Incorporating perceptual decision-making training into high-intensity interval training for Australian football umpires. *Journal of Sports Sciences, 37*(1), 29-35. https://doi.org/10.1080/02640414.2018.1480257

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174. https://doi.org/10.2307/2529310

Larkin, P., Berry, J., Dawson, B., & Lay, B. (2011). Perceptual and decision-making skills of Australian football umpires. *International Journal of Performance Analysis in Sport, 11*(3), 427-437. https://doi.org/10.1080/24748668.2011.11868562

Larkin, P., Mesagno, C., Berry, J., & Spittle, M. (2014). Development of a valid and reliable video-based decision-making test for Australian football umpires. *Journal of Science and Medicine in Sport, 17*(5), 552-555. https://doi.org/10.1016/j.jsams.2013.08.001

Larkin, P., Mesagno, C., Berry, J., & Spittle, M. (2018a). Exploration of the perceptual-cognitive processes that contribute to in-game decision-making of Australian football umpires. *International Journal of Sport and Exercise Psychology, 16*(2), 112-124. https://doi.org/10.1080/1612197X.2016.1167760

Larkin, P., Mesagno, C., Berry, J., Spittle, M., & Harvey, J. (2018b). Video-based training to improve perceptual-cognitive decision-making performance of Australian football umpires. *Journal of Sports Sciences, 36*(3), 239-246. https://doi.org/10.1080/02640414.2017.1298827

Larkin, P., Mesagno, C., Spittle, M., & Berry, J. (2015). An evaluation of video-based training programs for perceptual-cognitive skill development. A systematic review of current sport-based knowledge. *International Journal of Sport Psychology, 46*(6), 555-586.

Lorains, M., Ball, K., & MacMahon, C. (2013b). Expertise differences in a video decision-making task: Speed influences on performance. *Psychology of Sport and Exercise, 14*(2), 293-297. https://doi.org/10.1016/j.psychsport.2012.11.004

MacMahon, C., Helsen, W., Starkes, J., & Weston, M. (2007a). Decision-making skills and deliberate practice in elite association football referees. *Journal of Sports Sciences, 25*(1), 65-78. https://doi.org/10.1080/02640410600718640

Mann, D., Farrow, D., Shuttleworth, R., & Hopwood, M. (2009). The influence of viewing perspective on decision-making and visual search behaviour in an invasive sport. *International Journal of Sport Psychology, 40*(4), 546-564.

Panchuk, D., Klusemann, M. J., & Hadlow, S. M. (2018). Exploring the effectiveness of immersive video for training decision-making capability in elite, youth basketball players. *Frontiers in Psychology, 9*, 1-9. https://doi.org/10.3389/fpsyg.2018.02315

Paradis, K., Larkin, P., & O'Connor, D. (2016). The effects of physical exertion on decision-making performance of Australian football umpires. *Journal of Sports Sciences, 34*(16), 1535-1541. https://doi.org/10.1080/02640414.2015.1122205

Petit, J.-P., & Ripoll, H. (2008). Scene perception and decision making in sport simulation: a masked priming investigation. *International Journal of Sport Psychology, 39*(1), 1-19.

Plessner, H., & Betsch, T. (2001). Sequential effects in important referee decisions: The case of penalties in soccer. *Journal of Sport and Exercise Psychology, 23*(3), 254-259. https://doi.org/10.1123/jsep.23.3.254

Spitz, J., Put, K., Wagemans, J., Williams, A. M., & Helsen, W. F. (2017). Does slow motion impact on the perception of foul play in football? *European Journal of Sport Science, 17*(6), 748-756. https://doi.org/10.1080/17461391.2017.1304580

Vignais, N., Kulpa, R., Brault, S., Presse, D., & Bideau, B. (2015). Which technology to investigate visual perception in sport: Video vs. virtual reality. *Human Movement Science, 39*, 12-26. https://doi.org/10.1016/j.humov.2014.10.006

Williams, A. 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

Chapter 5: Transfer of 360° VR and Match Broadcast Video to Match Performance

This Chapter is presented in pre-publication format of a recent publication titled:

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. (2020). Transfer of 360° virtual reality and match broadcast video-based tests to on-field decision-making. *Science and Medicine in Football*. https://doi.org/10.1080/24733938.2020.1802506

VICTORIA UNIVERSITY

THE NEW WAY TO DO UNI

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book: **Kittel, A.**, Larkin, P., Elsworthy, N., & Spittle, M. (2020). Transfer of 360° virtual reality and match broadcast video-based tests to on-field decision-making. *Science and Medicine in Football*. https://doi.org/10.1080/24733938.2020.1802506

Surname: Kittel | First name: Aden

Institute: Institute for Health and Sport | Candidate's Contribution (%): 85

Status:

Accepted and in press: Date: July 2020

Published: Date:

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be included in the thesis as outlined in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted] | 12/8/2020

Signature **Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

PO Box 14428, Melbourne,
Vic 8001, Australia
+61 3 9919 6100

Victoria University ABN 83776954731
CRICOS Provider No. 00124K (Melbourne),
02475D (Sydney), RTO 3113

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location(s)**:

All electronic data will be stored on the Victoria University R Drive, which is a secure research storage drive.

Name(s) of Co-Author(s)	Contribution (%)	Nature of Contribution	Signature	Date
Paul Larkin	8	Reviewer, provided feedback, minor edits.	[redacted]	12/8/20
Nathan Elsworthy	3	Reviewer, provided feedback, minor edits.		12/8/20
Michael Spittle	4	Reviewer, provided feedback, minor edits.		12/8/20

Updated: September 2019

PO Box 14428, Melbourne,
Vic 8001, Australia
+61 3 9919 6100

Aden Kittel, Paul Larkin, Nathan Elsworthy & Michael Spittle (2021) Transfer of 360° virtual reality and match broadcast video-based tests to on-field decision-making, Science and Medicine in Football, 5:1, 79-86, DOI: 10.1080/24733938.2020.1802506

https://www.tandfonline.com/doi/abs/10.1080/24733938.2020.1802506

Chapter 6: Video-based Training Intervention

This chapter is presented in pre-publication format of a recent publication titled:

Kittel, A., Larkin, P., Elsworthy, N., Lindsay, R., & Spittle, M. (2020). Effectiveness of 360° virtual reality and match broadcast video to improve decision-making skill. *Science and Medicine in Football*. https://doi.org/10.1080/24733938.2020.1754449

VICTORIA UNIVERSITY

THE NEW WAY TO DO UNI

OFFICE FOR RESEARCH TRAINING, QUALITY AND INTEGRITY

DECLARATION OF CO-AUTHORSHIP AND CO-CONTRIBUTION: PAPERS INCORPORATED IN THESIS

This declaration is to be completed for each conjointly authored publication and placed at the beginning of the thesis chapter in which the publication appears.

1. PUBLICATION DETAILS (to be completed by the candidate)

Title of Paper/Journal/Book:	**Kittel, A.**, Larkin, P., Elsworthy, N., Lindsay, R., & Spittle, M. (2020). Effectiveness of 360° virtual reality and match broadcast video to improve decision-making skill. *Science and Medicine in Football*. https://doi.org/10.1080/24733938.2020.1754449
Surname: Kittel	First name: Aden
Institute: Institute for Health and Sport	Candidate's Contribution (%): 90

Status:

Accepted and in press:	Date:	April 2019
Published:	Date:	

2. CANDIDATE DECLARATION

I declare that the publication above meets the requirements to be included in the thesis as outlined in the HDR Policy and related Procedures – policy.vu.edu.au.

[redacted]	12/8/2020
Signature	**Date**

3. CO-AUTHOR(S) DECLARATION

In the case of the above publication, the following authors contributed to the work as follows:

The undersigned certify that:

1. They meet criteria for authorship in that they have participated in the conception, execution or interpretation of at least that part of the publication in their field of expertise;
2. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;

PO Box 14428, Melbourne,
Vic 8001, Australia
+61 3 9919 6100

3. There are no other authors of the publication according to these criteria;
4. Potential conflicts of interest have been disclosed to a) granting bodies, b) the editor or publisher of journals or other publications, and c) the head of the responsible academic unit; and
5. The original data will be held for at least five years from the date indicated below and is stored at the following **location(s)**:

All electronic data will be stored on the Victoria University R Drive, which is a secure research storage drive.

Name(s) of Co-Author(s)	Contribution (%)	Nature of Contribution	Signature	Date
Paul Larkin	5	Reviewer, provided feedback, statistical assistance, minor edits.	[redacted]	12/8/20
Nathan Elsworthy	2	Reviewer, provided feedback, minor edits.		12/8/20
Riki Lindsay	1	Reviewer, provided feedback, minor edits.		12/8/20
Michael Spittle	2	Reviewer, provided feedback, minor edits.		12/8/20

Updated: September 2019

6.1. Abstract

Video-based training is a commonly used method to develop decision-making skills in athletes and officials. This method typically uses match broadcast footage, yet technological advancements have made 360° Virtual Reality (360° VR) a possible effective tool to investigate. This study aimed to investigate the effectiveness of both 360° VR and match broadcast footage on improving decision-making. Amateur Australian football umpires (n = 32) participated in a randomised control trial. Participants completed a 5-week training intervention in either a 360° VR, match broadcast or control group. Decision-making was assessed at pre-training, 1-week post-training and 4-week retention points using reliable and valid 360° VR and match broadcast decision-making tests. Participants completed a short questionnaire detailing their perceptions of psychological fidelity, enjoyment, relevance, concentration and effort for each video mode. The 360° VR performed significantly better ($p < 0.05$) than the control group in the 360° VR retention test. No groups statistically improved over the intervention. Remaining pairwise comparisons for this test and the match broadcast test were not significantly different. 360° VR was rated significantly higher ($p < 0.05$) than match broadcast footage for psychological fidelity, enjoyment and relevance. 360° VR appears to be a beneficial training tool compared to a control, with stronger engagement from the participants than previously used match broadcast footage.

6.2. Introduction

Decision-making is a fundamental cognitive skill for successful sporting performance, defined as the ability to perceive and correctly interpret game-related information to select an appropriate sport-specific response (Baker, Côté, & Abernethy, 2003). Researchers have suggested decision-making skills are an effective means of identifying expert-novice differences in both athletes (Williams & Ericsson, 2005) and officials (Kittel, Larkin, Elsworthy, & Spittle, 2019d). To elicit expertise differences, it is important for the decision-making processes of the task to be similar to those experienced within competition, leading to stronger representativeness and ecological validity (Abernethy, Thomas, & Thomas, 1993; Farrow, Reid, Buszard, & Kovalchik, 2018). As decision-making is an important skill in players (Williams & Ericsson, 2005) and officials (Kittel, Larkin, Elsworthy, & Spittle, 2019b), it is imperative to understand the most efficient strategies to improve decision-making skills. Decision-making training tasks for athletes commonly include approaches such as small-sided games, which simulate the decision-making processes of match play in the training environment via self-guided discovery (O'Connor, Larkin, & Williams, 2017). This approach is an effective means of developing decision-making as it focuses holistically on the interaction between decision-making, tactical knowledge and skill execution, rather than isolating discrete performance components in conditions unlike performance (Light, Harvey, & Mouchet, 2014).

For sporting officials, it has been acknowledged the best approach to develop decision-making skill is through in-game practice (MacMahon, Helsen, Starkes, & Weston, 2007a). However, this approach is limited by the finite number of games, with inherent physical loads, to officiate (Elsworthy, Burke, Scott, Stevens, & Dascombe, 2014). While small-sided games are an easy and time efficient method to develop athletes' decision-making ability, from a sports official's decision-making development perspective, it is difficult and inefficient to organise and administer effective small-sided games for this cohort. A common approach that addresses these limitations and has effectively enhanced decision-making skill is video-based training, defined as a form of practice using video to present a situation requiring a perceptual-cognitive (i.e., decision-making) response from an individual (Larkin, Mesagno, Spittle, & Berry, 2015). Simulation practice such as video-based training recreates aspects of

a skill artificially, allowing the learner to practice without physically performing the skill (Spittle, 2013). As such, this form of practice is the most common modality for developing decision-making skills in sports officials.

Video-based training incorporating game footage (i.e., match broadcast) has been used with sporting officials such as soccer referees and demonstrated improvement of decision-making skill following a 7-week training intervention (Schweizer, Plessner, Kahlert, & Brand, 2011). Similarly in Australian football umpires, video-based training has been an effective means of developing decision-making skill, particularly in less experienced participants following a 12-week training intervention (Larkin, Mesagno, Berry, Spittle, & Harvey, 2018b). The most common modality for video-based training methods is match broadcast video. This method is however limited by presenting third-person perspective footage filmed from an elevated position in the grandstand (Craig, 2013). This viewpoint limits the representativeness of making decisions, as this is not the same perspective of an official in a game. As a key aim of video-based tasks is for individuals to make a decision using similar processes to what they would in a game situation (Farrow, 2013), there may be more suitable presentation methods for video-based training than existing match broadcast technology.

To overcome this limitation and increase the representativeness of existing video-based training tasks, researchers should optimise psychological fidelity, which refers to the extent a situation (i.e., decision-making in video-based training task) represents reality (i.e., decision-making in-game) (Alessi, 1988; Farrow, 2013). Virtual reality has been identified as a new video-based training method to increase the psychological fidelity of a task (Craig, 2013). As the videos in virtual reality present scenarios in the first-person viewing perspective, virtual reality may increase the representativeness of the task (Petit & Ripoll, 2008). In addition to the first-person perspective, virtual reality as experienced through a head mounted display (HMD) incorporates a stronger interaction with the environment and immersion than screen-based technology such as match broadcast video, as the vision automatically updates with the individual's head movements (Bird, 2020; Craig, 2013). In Australian football, 360° video viewed through a HMD has been shown to be a reliable and valid measure of umpires' decision-making skill (Kittel, Larkin, Elsworthy, & Spittle, 2019c). Importantly, the findings demonstrated 360°

VR provides a significantly stronger representation of match decision-making process than traditional match broadcast footage, inferring stronger psychological fidelity and ecological validity.

Further, 360° VR technology has also been examined as a video-based training tool for decision-making skill in basketball players (Panchuk, Klusemann, & Hadlow, 2018), which demonstrated large yet non-significant improvements. This suggests further research is required to determine the effectiveness of this technology for developing decision-making. Pagé, Bernier, and Trempe (2019) also developed a 360° VR intervention for basketball players, with a further comparison to two-dimensional screen-based video of the same scenario. Unlike Panchuk et al. (2018), decision-making changes were not assessed using a video-based (i.e., immersive 360° VR) test, but with an on-court transfer test. Decision-making was assessed through both trained and untrained plays in the on-court test. For trained plays, both intervention groups (i.e., immersive 360º VR and two-dimensional screen-based video) improved significantly compared to the control following the intervention. Improvements were only observed for the 360º VR group in untrained plays, with no such changes for the screen-based or control groups. This study provides promising findings for the use of 360° VR in developing decision-making skill in sport. To date, no studies have provided a comparison of the effectiveness of both 360° VR and traditional match broadcast video-based training methods.

Researchers in neuroscience suggest video-based training methods such as virtual reality are effective because it draws on a fundamental mechanism in the brain known as embodiment (Riva, Wiederhold, & Mantovani, 2019). Embodiment is described as the representation of the body in the brain, composed of vision, touch, proprioception, interoception, motor control and vestibular senses. Furthermore, embodiment refers to the integration of incoming sensory signals to the body, which the brain uses to generate a self-representation (Matamala-Gomez et al., 2019). Studies conducted in applied neuroscience have found that during simulated practice (e.g., virtual reality) individuals can experience a sense of embodiment (Slater et al., 2008). Being able to generate a greater sense of embodiment through simulated practice allows for the sensorimotor system to be engaged more fully, enhancing the potential to develop more desirable behavioural changes (Bohil, Alicea, & Biocca, 2011). Furthermore, not only does simulated practice allow for an individual to feel immersed in a virtual

world but it increases the ecological validity of the training being conducted. Ecological validity refers to the relationship between a cue (perceptual variable) to that of the performance environment (Araujo, Davids, & Passos, 2007). 360° VR presents more realistic and ecologically valid footage through first-person perspective and head movements than traditional 2D screen approaches, but does not allow physical interaction with the environment. 360° VR is a suitable training tool in populations where the video-based task can be 'read only' (i.e., does not require naturalistic interaction with the environment) such as American football quarterbacks and sports officials (Fadde & Zaichkowsky, 2018). Therefore from a theoretical perspective, modalities such as 360° VR may lead to stronger embodiment and potentially greater learning adaptations.

Therefore, the aim of this study was to determine the effectiveness of both 360° VR and traditional match broadcast video-based training methods in the development of decision-making skill. A secondary aim of this study was to examine the representativeness of each video mode through assessment of psychological fidelity and perception of deliberate practice.

6.3. Method

6.3.1. Participants

32 Australian football umpires participated in this research project, who regularly officiated in Division 1 senior Australian football grades in Metropolitan Melbourne competitions. The participants were randomly allocated to one of three groups; 360º VR training group (n = 12), match broadcast group (n = 10) and a control group (n = 10). The average age of the participants was 29 ± 13 years, with an average Australian football umpiring experience of 160 ± 117 games officiated. Ethics approval was granted by the lead institution's Human Research Ethics Committee (approval number: HRE18-181) and all participants provided informed consent prior to participation.

6.3.2. Measures

Two different video modalities were used for this study. Firstly, 360º video was captured from small-sided games of Australian football. This was filmed using a 360º video camera (360fly 4k Camera, 360fly, Canonsburg, USA) as part of normal training sessions of a high performance under 18 Australian football club [i.e., highest competition level of youth Australian football within the Australian Football talent pathway (Haycraft, Kovalchik, Pyne, & Robertson, 2017)]. The camera was positioned approximately 15 m to the side (i.e., perpendicular to the goals) of the contest, in a similar position an umpire would be in a game, to capture the first-person perspective. See a recent validation study (Kittel et al., 2019c) for a detailed explanation outlining the capture of this footage. Secondly, match broadcast (i.e., television) footage was obtained of elite Australian Football League (AFL) matches from the AFL Umpiring Department. 200 clips (not included in the test) from each video mode were presented to subject matter experts (i.e., AFL umpire coaches) for selection in the training battery. 125 clips were selected for each video mode. Final video clips were rendered in 4k video (3840 × 1920 pixels) for 360º VR and Full HD (1920 × 1080 pixels) for match broadcast video using ADOBE Premiere Pro CS4 video editing package. During the testing and training phases, all 360º VR footage was presented on an Oculus Go HMD (Oculus Go, Oculus, California, USA), and all match broadcast footage was presented on an iPad (Apple iPad Air 2, Apple Inc., Cupertino, USA). The two modalities have been validated in previous research (Kittel et al., 2019c).

To assess the effectiveness of the training intervention, this study utilised a reliable and valid video-based test from previous research (Kittel et al., 2019c). This test included the two video modes; 360º VR, and match broadcast, with each video mode including 25 decision-making scenarios. The two video modes were matched for number and type of decision-making incidents. For each video mode, participants completed three practice clips to familiarise themselves with the testing procedure. Each decision-making scenario had one correct decision determined by subject matter experts resulting in a maximum score of 25 for each video mode (Kittel et al., 2019c). Clip order was randomised between participants.

To assess the ecological validity of both video modes, all participants completed a short questionnaire outlining their perceptions of psychological fidelity and deliberate practice of each video mode immediately following the pre-test and prior to the training intervention, as this was their first exposure to the two video modes. Psychological fidelity of each video mode was assessed through a 10 cm visual analogue scale used in past research (Catteeuw, Gilis, Jaspers, Wagemans, & Helsen, 2010), which asked the participants to "make a dash on the line for how each method feels like your match decision-making process" from "nothing like match decision-making" to "exactly like match decision-making" (Kittel et al., 2019c). The distance of the dash marked on each visual analogue scale was measured in centimetres to one decimal place. To further understand the perceptions of deliberate practice toward each video mode, a questionnaire from past research (Catteeuw, Helsen, Gilis, & Wagemans, 2009; MacMahon et al., 2007a) was adapted. For the two video modes, participants were asked to provide a numerical rating from 1 (low) to 10 (high) indicating their perception of relevance, effort, enjoyment and concentration. For this study, the terms were defined as the following: relevance; "how appropriate is this for umpiring AFL?", effort; "how hard did you find the decision-making component of the task?", enjoyment; "would you happily use this as a coaching tool?" and concentration; "how much did you need to focus?".

6.3.3. Procedures

Upon providing informed consent, the participants completed the pre-test as outlined above to establish baseline decision-making scores. The presentation of each clip included a three second period outlining the clip number in white font on a black screen, followed by the first frame of the clip presented as a static image for two seconds to provide context (Kittel et al., 2019c). The video was then presented until a decision moment occurred (3 - 12 seconds duration), followed by the final frame presented as a static image for 0.5 seconds. Finally, a black screen with white font stating "make your decision" was presented for six seconds. The only instructions participants were given was to use the same decisional interpretation process they would in a game.

Following the pre-test, participants were randomly allocated into three groups. The pre-test occurred one week prior to the 5-week intervention period. The 360º VR group (n = 12) completed one video-based training session a week (10 min per session), with each session presenting 25 decision-making scenarios plus explicit feedback providing a detailed interpretation of the justification for the correct decision. The match broadcast group (n = 10) also completed one video-based training session a week (10 min per session) with 25 decision-making scenarios, all with feedback on the correct decision interpretation for each scenario. The two intervention groups had the same number of total decisions, decision types and feedback provided. The total decision load (125 clips for the intervention duration) was designed in light of previous research demonstrating a positive effect using 60 total clips for Australian football players (Lorains, Ball, & MacMahon, 2013) and 144 clips incorporating feedback on the correct decision for soccer referees (Schweizer et al., 2011). Two studies have examined the effectiveness of 360º video on improving decision-making skill, with 41 (Panchuk et al., 2018) and 120 original clips (Pagé et al., 2019) used as the training stimulus. Therefore, the authors concluded this would be an adequate total number of clips to elicit a performance improvement. The control group did not receive any additional video training between the pre- and post-testing sessions. Throughout the intervention, all groups continued with regular Australian football umpire training. This study was conducted in the off-season phase of competition, negating any impact from in-game practice.

One week following the training intervention, participants completed the post-test. This used the same testing protocol (i.e., same clips on the two video modes) as the pre-test to determine any decision-making performance improvements. Four weeks following the post-test, the participants completed a retention test, incorporating the same clips as the pre- and post-testing. No additional video was presented for any participants in this 4-week period. To minimise learning effects of the test, clip order and sequence of test administration (i.e., 360º VR or match broadcast test first) was randomised for all participants each testing session. Due to dropout, 25 participants completed the retention test; 360º VR group (n = 10), match broadcast group (n = 9), and control group (n = 6).

6.3.4. Statistical analysis

Data were assessed for normality using the Shapiro-Wilk test. Due to the non-normal distribution of data and the small sample size, non-parametric statistics were used to examine the effectiveness of the training intervention, both within- and between-groups. To analyse the differences between the independent groups over the course of the intervention, a Kruskal-Wallis test was conducted. Mann-Whitney U tests were conducted to examine pairwise group comparisons between the groups for each of the individual testing occasions. Friedman's ANOVA was used to assess within-group differences over the intervention. Parametric tests were used to examine the differences between video modes for the fidelity and deliberate practice questionnaire as these did not violate normality. A paired samples t-test analysed the differences between video modes for psychological fidelity (game likeness), relevance, effort, enjoyment and concentration. Statistical significance was set at $p = 0.05$. All tests were conducted using SPSS (IBM SPSS Statistics, SPSS Inc., USA). Effect size *r* was used, with values 0.1 - 0.3 representing a small effect, 0.3 - 0.5 medium effect, and greater than 0.5 being a large effect (Field, 2009).

6.4. Results

6.4.1. 360º VR test performance

For performance in the 360º VR test, the Kruskal-Wallis test identified no significant differences between groups at the pre, $H(2) = 0.84$, $p = 0.66$ and post testing sessions, $H(2) = 4.73$, $p = 0.09$. There was a significant difference observed at the retention test, $H(2) = 7.05$, $p = 0.05$. Post-hoc analysis indicated the 360º VR scored significantly higher (see Figure 6.1) than the control group at the retention ($p = 0.03$, $r = 0.54$; *large*) tests. There were no significant differences between the match broadcast and control groups at the retention test ($p = 0.11$, $r = 0.36$; *medium*). Similarly, there were no significant differences between the 360º VR and match broadcast groups in the retention test ($p = 0.16$, $r = 0.32$; *medium*). These results are illustrated in Figure 6.1.

Friedman's ANOVA reported no differences within groups across the intervention as assessed in the 360º VR test for the 360º VR; $\chi 2(2) = 2.14$, $p = 0.35$, match broadcast; $\chi 2(2) = 3.18$, $p = 0.15$, and control; $\chi 2(2) = 3.60$, $p = 0.16$ groups.

6.4.2. Match broadcast test performance

For the match broadcast test, there were no significant differences at the pre, $H(2) = 0.24$, $p = 0.89$, post, $H(2) = 1.88$, $p = 0.39$ and retention, $H(2) = 0.42$, $p = 0.88$) testing sessions. These results are illustrated in Figure 6.1.

There were no differences within groups across the intervention as assessed in the match broadcast test for the 360º VR; $\chi 2(2) = 2.14$, $p = 0.35$, match broadcast; $\chi 2(2) = 1.07$, $p = 0.59$, and control; $\chi 2(2) = 0.23$, $p = 0.89$ groups.

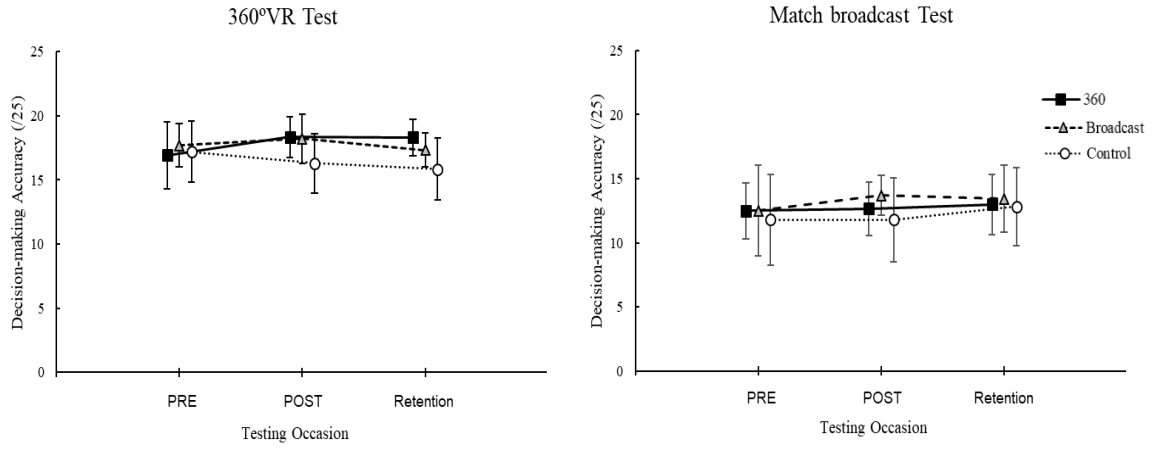

Figure 6.1: Decision-making performance changes following the training intervention.

6.4.3. Perceptions of fidelity & deliberate practice questionnaire

The participants rated the 360º VR video condition to have significantly higher psychological fidelity (7.02 ± 1.69) compared to the match broadcast condition (4.76 ± 2.40), $t(26) = 4.36$, $p = 0.01$, $r = 0.37$. 360º VR (8.11 ± 1.42) was considered to have greater relevance to officiating Australian

football than the match broadcast (6.75 ± 1.84), $t(27) = 5.26$, $p = 0.01$, $r = 0.71$. Similarly, participants rated the 360º VR (8.93 ± 1.09) to have greater enjoyment than match broadcast (7.68 ± 1.72), $t(27) = 3.49$, $p = 0.01$, $r = 0.56$. There were no significant differences for effort between 360º VR (5.93 ± 2.09) and match broadcast (6.07 ± 2.09), $t(27) = -0.46$, $p = 0.65$, $r = 0.09$. There were also no differences in concentration for 360º VR (7.93 ± 1.52) and match broadcast (7.37 ± 1.57), $t(26) = 1.58$, $p = 0.65$, $r = 0.29$. These results are illustrated in Figure 6.2.

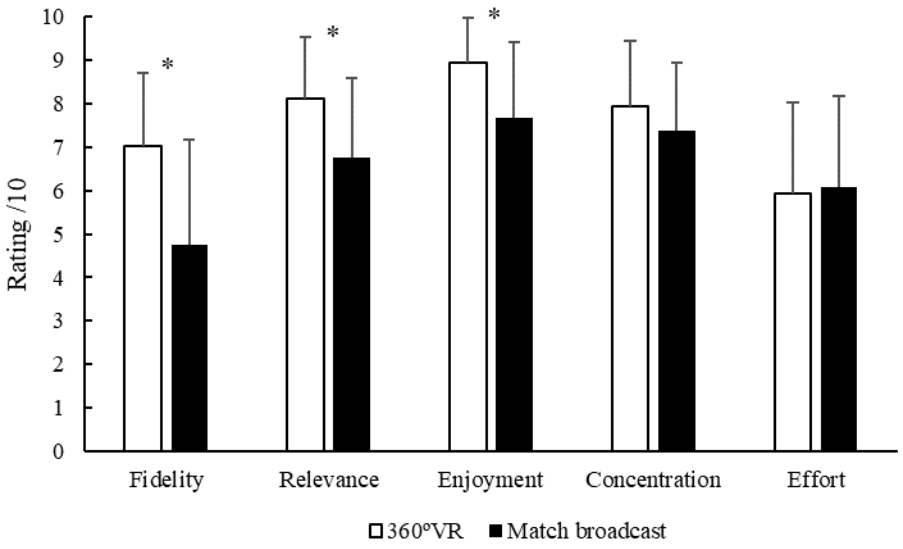

Figure 6.2: Perceptions of psychological fidelity and deliberate practice.

6.5. Discussion

The aim of this study was to determine the effectiveness of both 360º VR and match broadcast video-based training on the development of decision-making skill in Australian football umpires. Further, this study also aimed to determine the ecological validity of each video mode as assessed through psychological fidelity ratings and perceptions of deliberate practice. This was the first study to compare 360º VR and the previously investigated match broadcast video modality. There were no significant changes across the intervention for each group. The results demonstrate the 360º VR group outperformed the control group in the 360ºV R retention test following the training intervention. There

were no significant differences observed for the remaining between-groups comparisons. The perceptions of the participants indicate 360º VR is a more game-like (as assessed through psychological fidelity rating), enjoyable and relevant coaching tool for Australian football umpires than match broadcast footage. There was no difference between video modes for concentration and effort.

6.5.1. 360º VR and control group differences

Given the proposed benefits of virtual reality as a perceptual-cognitive tool (Craig, 2013), this study investigated whether 360º VR would be an effective video-based training modality. Although the 360º VR group did not statistically improve following the intervention, this group scored significantly higher than the control (i.e., no video-based training) in the 360º VR retention test. Two studies have examined the effectiveness of 360º VR in improving sport decision-making skill, both within basketball players. Pagé et al. (2019) demonstrated a training benefit of 360º VR and first-person screen-based intervention groups, with 360º VR being more beneficial for untrained plays. Panchuk et al. (2018) reported improvements in the female control and intervention groups, but no change in male groups, which could be due to low sample size, training dose size to cause an effect and lack of a retention test to determine any long term effects. Retention tests are an important component of an intervention to determine performance learning changes (Lorains et al., 2013). Given the differences of the 360º VR group compared to the control in the present study, this technology may be a beneficial video-based training tool for practitioners, when compared to no additional video-based training. A greater sense of embodiment in a training tool increases the chances of long term behavioural changes (Bohil et al., 2011), which may explain why the 360º VR group, not the match broadcast group, was significantly higher than control at the retention test.

The findings of the current study suggest this novel video-based training method may have similar effects to existing video-based approaches demonstrating an improvement in decision-making in officials (Schweizer et al., 2011) and players (Lorains et al., 2013). The effectiveness of the 360º VR group compared to the control could be attributed to the stronger representativeness of this method

(Pinder, Davids, Renshaw, & Araújo, 2011), where the first-person footage provides stronger visual correspondence to information perceived in a game. Stronger visual correspondence of HMD rather than screen-based technology is an important consideration of the modified perceptual training framework (Hadlow, Panchuk, Mann, Portus, & Abernethy, 2018). This framework suggests the stronger visual correspondence, as afforded by the first-person perspective and head movements of the HMD technology theoretically, has more positive training effects than screen-based technology, which may limit immersion (Hadlow et al., 2018; Pagé et al., 2019). The behavioural correspondence is also stronger in the 360º VR method, where participants provided a verbal and motor response (i.e., free kick signal) as they would in a game and changes in viewing perspective were caused by participants' head movements. Stronger visual and behavioural correspondence are key aspects of representative design (Hadlow et al., 2018; Pinder et al., 2011) and could have contributed to the positive learning effect experienced by the 360º VR group in comparison to no training.

6.5.2. Differences of match broadcast training to 360º VR and control groups

Despite the difference observed between the 360º VR group and control, there were no significant differences reported for the match broadcast group in comparison to the control or 360º VR groups. A potential explanation for this result could be the number of clips used throughout the intervention of each group, which is a common discussion point with a large variance in methods applied between studies (Larkin, Mesagno, Spittle, & Berry, 2015). As highlighted in the methods section of this study, the duration was designed with a similar load to past video-based training studies presenting match broadcast footage (Lorains et al., 2013; Schweizer et al., 2011). In addition, the load of the current program (125 clips) was greater than previous 360º VR video-based training presenting 41 clips (Panchuk et al., 2018). Feedback was incorporated in this study in a similar manner to past research (Schweizer et al., 2011). For consistency, the clip type and number presented in the two intervention groups were the same to determine potential differences. Feedback was utilised to increase the learning effects of this intervention, as there were a finite amount of 360º VR clips available. As such, an implicit learning approach was not feasible within this study, as per previous video-based

training in Australian football umpires using 1040 clips (Larkin et al., 2018b). Participants currently train with match broadcast video and this could have resulted in a ceiling effect, where no significant changes were observed in this group.

Research has suggested a comparison of 360º VR and other video-based training modalities (e.g., match broadcast) (Panchuk et al., 2018), yet no differences were observed between these in the current study. Researchers have found athletes' decision-making improved following training using 360º VR scenarios of small-sided games compared to screen-based presentations (i.e., not immersive HMD) of the same scenarios (Pagé et al., 2019). The current study provided a comparison of 360º VR to match broadcast, due to previous research using structured match broadcast video-based training (Larkin et al., 2018b) and also match broadcast footage being used by coaches as a training tool. The results of this study suggest promising results for the use of 360º VR, as there was a statistical performance difference to the control group at retention testing, in addition to greater fidelity, relevance and enjoyment. It should still be noted there were no statistically significant differences between the two intervention groups (i.e., 360º VR and match broadcast). There is however a stronger level of embodiment afforded through using HMD rather than screen-based technology (Kilteni, Groten, & Slater, 2012), which is due to the stereovision/binocular vision (Faure, Limballe, Bideau, & Kulpa, 2020). From a neuroscientific perspective, embodiment allows the sensorimotor system to have greater engagement and maximise potential for behavioural changes (Bohil et al., 2011). The stronger embodiment through 360º VR HMD technology may contribute to the difference between 360º VR and control groups at retention test.

As no significant differences were observed between the two intervention groups, this may be due to the limited number of clips and sample size of the groups, despite the theoretical strengths of HMD versus screen-based technologies. Although participants have not previously used 360º VR as a video-based training tool, the number of clips used may explain why no changes were observed. Similar to physical training, a limited number of training stimuli presented in video-based training may limit the developmental changes over the intervention, by not allowing overload or progression of the skill (Farrow & Robertson, 2016). This consideration is pertinent to our hypothesis suggesting 360º VR to

be a superior training tool, considering the literature base outlining the effectiveness of HMD video-based training rather than existing methods (Bird, 2020; Craig, 2013; Fadde & Zaichkowsky, 2018). The authors must also acknowledge the limited group size may impact the statistical power of the study. As discussed by Panchuk et al. (2018), studies with smaller sample sizes remain useful, as some populations such as sporting officials are limited in the pool of participants to be involved.

6.5.3. Perceptions of fidelity and deliberate practice

The participants of this study had positive perceptions of 360º VR as a training tool compared to traditionally used match broadcast footage, as evident by the higher ratings of psychological fidelity, enjoyment and relevance in comparison to match broadcast footage. No differences were observed for concentration and effort. Deliberate practice is defined as being high in relevance, enjoyment, concentration and effort (Starkes et al., 1996). Studies in soccer referees have determined the most appropriate form of deliberate practice to be officiating matches, with this activity receiving higher ratings of relevance and enjoyment compared to all other training activities (Catteeuw et al., 2009; MacMahon et al., 2007a). Although officiating matches is the optimal method for developing skill, the perceptions of the participants in this study suggest that 360º VR may be a more appropriate form of deliberate practice for this cohort to complement in-game practice. Participants within this study attributed the stronger enjoyment to the novelty of experiencing 360º VR technology and the stronger game likeness of the decision-making. The stronger visual (i.e., first-person viewpoint) and behavioural correspondence (i.e., head movements) led to a more representative, game-like experience (Pinder et al., 2011), potentially through stronger embodiment (Riva et al., 2019). As umpires cannot routinely participate in small-sided games as athletes do for a representative training experience (O'Connor et al., 2017), this technology could provide an effective, enjoyable and relevant training tool for officials. Despite further research required to understand the effectiveness of 360º VR compared to match broadcast for decision-making development, the findings suggest athletes or officials may have greater satisfaction using 360º VR technology given the higher enjoyment and relevance ratings.

6.5.4. Study limitations

There are several limitations in this study that must be highlighted. Firstly, there were a limited number of clips used in the training intervention compared to past research (Larkin et al., 2018b). This study presented more training clips than a recent study examining the effectiveness of 360º VR video in developing sport-specific decision-making skill (Panchuk et al., 2018). Future studies should investigate a longer training duration using this technology to determine long-term effects. Transfer of training adaptations is an important consideration for research investigating the effectiveness of off-field decision-making training (Larkin et al., 2015) and testing approaches (Kittel et al., 2019d). These data were not available in this study, given the amateur cohort where there is less consistency in coaching in relation to more elite cohorts. The investigation of 360º VR in developing decision-making in more skilled cohorts needs to be considered in further research, as does the transfer of learning to on-field performance. Although numerous studies have investigated video-based training using amateur participants (Kittel, Elsworthy, & Spittle, 2019a; Larkin et al., 2017), future studies should investigate such approaches with other groups such as junior, sub-elite and elite. With the advancement of technology, future studies will implement more advanced versions of 360º VR and other technologies such as first-person match footage to assist with decision-making skill development.

6.5.5. Conclusion

To conclude, the 360º VR and match broadcast groups did not statistically improve over the course of the intervention. The 360º VR group scored higher than the control in the 360º VR retention test, suggesting this may be an appropriate training tool. No such differences between 360º VR and match broadcast intervention groups were observed. The findings suggest match broadcast training does not appear to be more beneficial in relation to normal training, which may be due to the low number of intervention clips unable to elicit an effect. This study has several practical implications for deliberate practice as the participants considered 360º VR to be a more relevant and enjoyable training tool than match broadcast footage, with no greater concentration or effort required. Further studies should

examine longer training programs of both video modalities to further elicit differences between the two video modes. Although there are promising findings for the use of 360º VR as a decision-making training tool, future research should also explore the potential use of other technologies to supplement in-game decision-making practice.

6.5.6. Practical implications

- A 360º VR training intervention demonstrates some improvements in decision-making skill in comparison to no video-based training (i.e., a control group).
- A five-week match broadcast training intervention does not elicit a significant change in decision-making skill.
- 360º VR is a more game-like decision-making training tool, with participants perceiving this video mode to have stronger relevance and enjoyment than traditional match broadcast footage.

6.6. References

Abernethy, B., Thomas, K. T., & Thomas, J. T. (1993). Strategies for improving understanding of motor expertise [or mistakes we have made and things we have learned!!]. *Advances in Psychology, 102*, 317-356. https://doi.org/10.1016/S0166-4115(08)61478-8

Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-based Instruction, 15*, 40-47.

Araujo, D., Davids, K., & Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: Comment on Rogers, Kadar, and Costall (2005). *Ecological Psychology, 19*(1), 69-78. https://doi.org/10.1080/10407410709336951

Baker, J., Côté, J., & Abernethy, B. (2003). Learning from the experts: Practice activities of expert decision makers in sport. *Research Quarterly for Exercise and Sport, 74*(3), 342-347. https://doi.org/10.1080/02701367.2003.10609101

Bird, J. M. (2020). The use of virtual reality head-mounted displays within applied sport psychology. *Journal of Sport Psychology in Action, 11*(2), 115-128. https://doi.org/10.1080/21520704.2018.1563573

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature reviews neuroscience, 12*(12), 752-762. https://doi.org/10.1038/nrn3122

Catteeuw, P., Gilis, B., Jaspers, A., Wagemans, J., & Helsen, W. (2010b). Training of perceptual-cognitive skills in offside decision making. *Journal of Sport and Exercise Psychology, 32*(6), 845-861. https://doi.org/10.1123/jsep.32.6.845

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

Craig, C. (2013). Understanding perception and action in sport: how can virtual reality technology help? *Sports Technology, 6*(4), 161-169. https://doi.org/10.1080/19346182.2013.855224

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

Fadde, P. J., & Zaichkowsky, L. (2018). Training perceptual-cognitive skills in sports using technology. *Journal of Sport Psychology in Action, 9*(4), 239-248. https://doi.org/10.1080/21520704.2018.1509162

Farrow, D. (2013). Practice-enhancing technology: a review of perceptual training applications in sport. *Sports Technology, 6*(4), 170-176. https://doi.org/10.1080/19346182.2013.875031

Farrow, D., Reid, M., Buszard, T., & Kovalchik, S. (2018). Charting the development of sport expertise: challenges and opportunities. *International Review of Sport and Exercise Psychology, 11*(1), 238-257. https://doi.org/10.1080/1750984X.2017.1290817

Farrow, D., & Robertson, S. (2016). Development of a skill acquisition periodisation framework for high-performance sport. *Sports Medicine, 47*(6), 1043-1054. https://doi.org/10.1007/s40279-016-0646-2

Faure, C., Limballe, A., Bideau, B., & Kulpa, R. (2020). Virtual reality to assess and train team ball sports performance: A scoping review. *Journal of Sports Sciences, 38*(2), 192-205. https://doi.org/10.1080/02640414.2019.1689807

Field, A. (2009). *Discovering statistics using SPSS*: Sage Publications.

Hadlow, S. M., Panchuk, D., Mann, D. L., Portus, M. R., & Abernethy, B. (2018). Modified perceptual training in sport: A new classification framework. *Journal of Science and Medicine in Sport, 21*(9), 950-958. https://doi.org/10.1016/j.jsams.2018.01.011

Haycraft, J. A., Kovalchik, S., Pyne, D. B., & Robertson, S. (2017). Physical characteristics of players within the Australian Football League participation pathways: a systematic review. *Sports Medicine-Open, 3*(1), 1-16. https://doi.org/10.1186/s40798-017-0109-9

Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and virtual environments, 21*(4), 373-387. https://doi.org/10.1162/PRES_a_00124

Kittel, A., Elsworthy, N., & Spittle, M. (2019a). Incorporating perceptual decision-making training into high-intensity interval training for Australian football umpires. *Journal of Sports Sciences, 37*(1), 29-35. https://doi.org/10.1080/02640414.2018.1480257

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. (2019b). Identification of key performance characteristics of elite Australian football umpires. *International Journal of Sports Science and Coaching, 14*(4), 490-497. https://doi.org/10.1177%2F1747954119845071

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. (2019c). Using 360° virtual reality as a decision-making assessment tool in sport. *Journal of Science and Medicine in Sport, 22*(9), 1049-1053. https://doi.org/10.1016/j.jsams.2019.03.012

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

Larkin, P., Mesagno, C., Berry, J., Spittle, M., & Harvey, J. (2018b). Video-based training to improve perceptual-cognitive decision-making performance of Australian football umpires. *Journal of Sports Sciences, 36*(3), 239-246. https://doi.org/10.1080/02640414.2017.1298827

Larkin, P., Mesagno, C., Spittle, M., & Berry, J. (2015). An evaluation of video-based training programs for perceptual-cognitive skill development. A systematic review of current sport-based knowledge. *International Journal of Sport Psychology, 46*(6), 555-586.

Light, R. L., Harvey, S., & Mouchet, A. (2014). Improving 'at-action'decision-making in team sports through a holistic coaching approach. *Sport, Education and Society, 19*(3), 258-275. https://doi.org/10.1080/13573322.2012.665803

Lorains, M., Ball, K., & MacMahon, C. (2013a). An above real time training intervention for sport decision making. *Psychology of Sport and Exercise, 14*(5), 670-674. https://doi.org/10.1016/j.psychsport.2013.05.005

MacMahon, C., Helsen, W., Starkes, J., & Weston, M. (2007a). Decision-making skills and deliberate practice in elite association football referees. *Journal of Sports Sciences, 25*(1), 65-78. https://doi.org/10.1080/02640410600718640

Matamala-Gomez, M., Donegan, T., Bottiroli, S., Sandrini, G., Sanchez-Vives, M. V., & Tassorelli, C. (2019). Immersive virtual reality and virtual embodiment for pain relief. *Frontiers in Human Neuroscience, 13*, 279. https://doi.org/10.3389/fnhum.2019.00279

O'Connor, D., Larkin, P., & Williams, A. M. (2017). What learning environments help improve decision-making? *Physical Education and Sport Pedagogy, 22*(6), 647-660. https://doi.org/10.1080/17408989.2017.1294678

Pagé, C., Bernier, P.-M., & Trempe, M. (2019). Using video simulations and virtual reality to improve decision-making skills in basketball. *Journal of Sports Sciences, 37*(21), 2403-2410. https://doi.org/10.1080/02640414.2019.1638193

Panchuk, D., Klusemann, M. J., & Hadlow, S. M. (2018). Exploring the effectiveness of immersive video for training decision-making capability in elite, youth basketball players. *Frontiers in Psychology, 9*, 1-9. https://doi.org/10.3389/fpsyg.2018.02315

Petit, J.-P., & Ripoll, H. (2008). Scene perception and decision making in sport simulation: a masked priming investigation. *International Journal of Sport Psychology, 39*(1), 1-19.

Pinder, R. A., Davids, K. W., Renshaw, I., & Araújo, D. (2011). Representative learning design and functionality of research and practice in sport. *Journal of Sport and Exercise Psychology, 33*(1), 146-155. https://doi.org/10.1123/jsep.33.1.146

Riva, G., Wiederhold, B. K., & Mantovani, F. (2019). Neuroscience of virtual reality: From virtual exposure to embodied medicine. *Cyberpsychology, Behavior, and Social Networking, 22*(1), 82-96. https://doi.org/10.1089/cyber.2017.29099.gri

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

Spittle, M. (2013). *Motor learning and skill acquisition: applications for physical education and sport*: Palgrave Macmillan Australia.

Starkes, J. L., Deakin, J. M., Allard, F., Hodges, N. J., & Hayes, A. (1996). Deliberate practice in sports: What is it anyway? In K. A. Ericsson (Ed.), *The Road to Excellence: The Acquisition*

of Expert Performance in the Arts, Sciences, Sports and Games (pp. 81-106). Mahwah, NJ: Lawrence Erlbaum.

Williams, A. 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

Chapter 7: Summary, Implications and Conclusion

This chapter aims to summarise the findings presented in this thesis. Firstly, brief summaries are provided of each study included in this thesis, followed by a discussion which aims to link together each of the studies presented, with reference to past research studies in this field. Practical implications of the findings herein are outlined to assist practitioners in officiating populations. Furthermore, the limitations of the research conducted are acknowledged, followed by recommendations for future research in this area based on inherent limitations of the current studies. Finally, concluding remarks are outlined to summarise this doctoral investigation.

7.1. Study 1 summary

Study 1 of this thesis was a systematic review summarising the existing literature base of video-based testing in sports officials. Decision-making is regarded as the most important skill for sports officials, with video-based approaches being the most common method used to reliably assess this skill. Researchers have shown significant interest in the use of this modality, leading to this study being undertaken. Following PRISMA guidelines, 27 articles were identified from January 2000 to January 2018. The majority of studies identified (16 of 27) investigated soccer officials, suggesting video-based testing should be examined in other populations. This systematic review highlighted several key findings and subsequent recommendations for this field of research. Video-based testing is often able to distinguish between skill levels of sports officials, thus indicating strong construct validity. Although the validity of video-based testing appears to be sound, reliability of tests is rarely reported, which is essential to establish the reproducibility of the tool, especially as these tools are often used to examine the effectiveness of decision-making training interventions. Studies should also report the accuracy of decisions made in video-based tests, rather than solely the number of decisions. Future researchers should analyse the transfer of decision-making accuracy in video-based tests to in-game accuracy. These data are needed to distinguish within population groups such as elite participants and further validate these tools to be used for selection and talent identification. Finally, there is a need to

investigate novel presentation modes such as virtual reality to enhance the representativeness of video-based testing. This study was essential in summarising the current knowledge base of video-based testing in sports officials.

7.2. Study 2 summary

Study 1 of this thesis highlighted several limitations of existing video-based testing studies in sports officials. For example, there is a need to establish the reliability and validity of video-based testing tools, as recommended by Larkin et al. (2015). A key point identified was that many video-based testing studies do not establish both the validity and reliability of the video tool being used, which is a significant consideration. Furthermore, additional video modes such as virtual reality warrant investigation as more representative and ecologically valid presentation methods. In consideration of these findings, Study 2 aimed to establish the reliability, construct validity and ecological validity of two different video modes, including; 360° VR and match broadcast footage (normally used in past research). This study followed a similar design to past research (Larkin et al., 2014a), with test re-test reliability analysed through two testing sessions of each video-based test. Construct validity was assessed through skill-based differences of amateur and elite participants. Ecological validity was determined by using visual analogue scales to determine the game-likeness of each video mode. ICC scores indicated both video modes had good test re-test reliability (ICC = 0.89 for both video modes), inferring strong reproducibility to assess decision-making changes over time. The two video modes had strong construct validity, where elite participants scored significantly higher in decision-making accuracy compared to their amateur counterparts. There was stronger ecological validity in the 360° VR condition, with participants rating this to have more game-likeness in relation to their match decision-making processes than match broadcast footage. Importantly, this is the first study to establish the reliability and validity of 360° VR as a decision-making assessment tool in sport, suggesting this may be an appropriate testing tool for talent benchmarking or evaluating the effectiveness of future

training interventions. This study highlighted stronger ecological validity and game-likeness afforded in 360° VR than match broadcast video, suggesting that 360° VR could be a stronger training tool.

7.3. Study 3 summary

Study 3 aimed to determine the relationship between decision-making accuracy in the two reliable and valid video-based tests of Study 2 to accuracy in competitive matches. Study 1 identified that video-based testing in sporting officials rarely examines the relationship between off-field (i.e., video-based test) decision-making performance to that of in-game, which is also a limited area of research in playing populations. Study 3 attempted to address this gap in the literature. Elite officials (n = 21) were recruited for this research, given their in-game performance is measured in a more robust and reliable procedure than lower performance levels. All participants completed the reliable and valid tests established in Study 2 to ascertain decision-making performance in an off-field setting. On-field decision-making was calculated from four matches at AFL level, where coaches provide a detailed assessment of decision-making performance. There was no significant relationship (r = 0.44) between decision-making accuracy in either video-based test to that of in-game. This result may be due to a number of factors such as the elite officials in this study watching the 360° VR where youth participants were filmed. 360° VR may limit the amount of contextual information that would be adjudicated in-game, which is required to differentiate between participants. Alternatively, the match broadcast video-based test presents a different viewpoint to the first-person view in-game. This limits the similarity of perceptual information received by the participant, thereby diminishing expertise differences. This study provides recommendations to make tasks more representative of the competitive environment by including more constraints that are in competition. Although this study could not identify a decision-making tool which mirrors the competitive environment, this research addressed a significant gap outlined in Study 1. Future studies could examine whether first-person game video presents a stronger representation of in-game decision-making.

7.4. Study 4 summary

Study 4 aimed to investigate the effectiveness of 360° VR and match broadcast in a video-based training intervention. Participants included 32 amateur Australian football umpires who completed a five-week video-based training intervention with the duration of the intervention based on previous research (Lorains et al., 2013a). The participants were randomly allocated into three groups, including; 360° VR (n = 12), match broadcast video (n = 10), and control (i.e., no training; n = 10). This was the first empirical study to use both 360° VR and match broadcast video in a matched intervention, where both video modes are matched for consistent scenarios. The reliable and valid tests of Study 2 were used to assess decision-making one week before (pre-test) and one week after (post-test) the five-week intervention, in addition to a retention test four weeks after the post-test. There were no significant changes for any groups over the course of the training intervention. While no significant between-group changes were evident for the post-test, there was a significant difference in decision-making between the 360° VR and control groups in the 360° VR retention test. This finding suggests 360° VR may be more beneficial than no training for enhancing decision-making performance, and supports the notion that virtual reality such as 360° VR approaches can lead to long-term behavioural changes. This result may be due to the greater immersion and embodiment of the 360° VR technology viewed through a HMD. In addition to being rated to have stronger game-likeness similar to Study 2, 360° VR was rated to have greater enjoyment and relevance than match broadcast video. This carries implications for deliberate practice approaches in sport, given practice activities should have high levels of enjoyment, relevance, concentration, and effort. Based on these findings, it not clear whether 360° VR is a more effective training tool than match broadcast, though there are important results which suggest this technology should be investigated further.

7.5. Discussion

This thesis initially aimed to summarise the existing literature on video-based testing in sports officials and highlight areas where this research can progress. The findings of Study 1 led to Study 2 establishing the reliability of two video-based tests, followed by Study 3 examining the level of transfer from video-based testing to on-field performance. Given the stronger ecological validity of 360° VR than match broadcast video, these two video modes were assessed as training tools in Study 4. Given the importance of decision-making to officiating performance (Kittel et al., 2019b; Morris & O'Connor, 2016), further understanding of the methods to investigate this skill is required, particularly in off-field testing and training approaches. Off-field decision-making testing is used in the literature for a number of reasons, such as identifying expert-based differences or to understand decision-making behaviour. This may include the effect of crowd noise (Nevill, Balmer, & Williams, 2002), biases (Krenn, 2014) and effects of physical exertion (Paradis et al., 2016). Video-based training, on the other hand, aims to improve the fundamental skill of decision-making, given the finite number of matches available to participate. The limited number of matches limits the amount of deliberate practice time. Therefore, alternative modes of training to that of on-field practice are required, with the most common example being video-based training.

Study 1 identified several limitations of previous research using video-based tests to assess decision-making skill in controlled environments. A key potential limitation of the majority of studies reviewed was the use of a match broadcast perspective. Theoretically, match broadcast video limits the perceptual information received, which in turn decreases ecological validity (Araujo et al., 2007). The game-likeness scale of Study 2 quantified the decreased ecological validity of the existing match broadcast method. For example, given match broadcast video is filmed from a fixed and elevated position in the grandstand, this footage does not incorporate the in-game perspective of players or officials (Craig, 2013). This perspective would limit the ability for participants to make decisions with similar perceptual processes to game situations (Farrow, 2013). In addition, match broadcast vision presented on a screen does not update with head movements, which may limit the representative nature and immersion of the decision-making task (Bird, 2020). With Study 1 identifying the use of match

broadcast video being the predominant presentation mode in sports officials' video-based testing, further investigation of other technologies is necessary.

360° VR was identified as a suitable technology which overcomes some of the limitations of match broadcast video described above, subsequently leading to the investigation of 360° VR in Studies 2, 3, and 4. This technology has been labelled as a suitable 'middle ground' between virtual environments and screen-based approaches such as match broadcast (Fadde & Zaichkowsky, 2018). This is due to the real world footage presented in 360° VR, integrated within a more representative presentation method (i.e., the HMD) where the vision automatically updates with the participants' head movements, thus increasing presence (Bird, 2020; Slater, 2018). Theoretically, technologies such as 360° VR lead to a stronger sense of embodiment and presence, which may subsequently lead to more desirable behaviour changes such as improved decision-making over a longer period of time (Bohil et al., 2011). Although studies have suggested 360° VR is a more representative and ecologically valid approach (Fadde & Zaichkowsky, 2018; Hadlow et al., 2018), there is limited research quantifying this. Therefore, Study 2 adopted a game-likeness scale from previous research (Catteeuw et al., 2010b) where participants rated 360° VR and match broadcast video approaches to the extent each represented the in-game decision-making process. This was to quantify the ecological validity of each approach (i.e., how similar decision-making processes are to in-game) (Araujo et al., 2007), with 360° VR displaying significantly higher ratings of game-likeness, particularly in amateur participants. This has also been referred to throughout this thesis as a rating of psychological fidelity, which similar to ecological validity, is how comparable to competition the simulation is perceived to be (Lorains et al., 2013a; Stoffregen et al., 2003). As it is important for players also to use ecologically valid and representative training approaches, technologies such as 360° VR may be appropriate for athletic populations training perceptual-cognitive skills that do not require a movement response.

Creating tasks with strong fidelity (i.e., how much it simulates competition/reality) is an important consideration for representative design (Alessi, 1988; Farrow, 2013). Tasks with stronger representativeness and/or ecological validity will be more likely to simulate expertise-based differences

in testing with transfer to in-game performance, and lead to greater training adaptations (Farrow et al., 2018; Hadlow et al., 2018). For training adaptations to occur, representative learning design suggests that practice tasks must represent competition with similar information and cues perceived to that of competition, accompanied by a sport-specific response (Hadlow et al., 2018; Pinder et al., 2011). The importance of perceptual cues in this framework is similar to the concept of ecological validity described above (Araujo et al., 2007). For a population such as sporting officials who do not require a sport-specific action such as a pass or intercept, 'read only' technologies such as 360° VR (Fadde & Zaichkowsky, 2018) incorporate the necessary game-like perceptual cues outlined in the representative learning design and ecological validity frameworks. Therefore, this technology warranted investigation for off-field testing and training purposes.

From the systematic review conducted in Study 1, there were a number of video-based testing limitations identified beyond the use of match broadcast video, and the subsequent rationale for 360° VR described above. For instance, it was evident that video-based tests need to investigate officials outside of soccer referees, establish reliability in addition to construct validity, and examine the level of transfer to match decision-making performance. Studies 2 and 3 aimed to address the limitations of existing video-based testing methods in Australian football officials, particularly due to the importance of this skill to overall performance (Kittel et al., 2019b). Firstly, Study 2 examined the reliability, construct validity and ecological validity (game-likeness) of both 360° VR and match broadcast video modes. In addition to the ecological validity outlined above, both video modes demonstrated strong construct validity by differentiating between skill levels similar to video-based tests in different sporting officials such as Australian football umpires (Larkin et al., 2014a) and soccer referees (Spitz et al., 2016). Beyond officials, video-based tests also demonstrated construct validity in playing populations (Lorains et al., 2013b). Further, both video modes resulted in strong levels of reliability as per previous research (Larkin et al., 2014a; Spitz et al., 2017). Given the high reliability, it was concluded these video-based tests could be used to assess performance changes following an intervention, such as in Study 4.

Secondly, Study 3 addressed the gap identified in Study 1 where only one study (Nazarudin et al., 2015) has provided a comparison of decision-making performance in video-based tests to that of on-field decision-making performance. This study assessed the transfer between match decision-making performance and decision-making in two video-based tests. It was hypothesised the underlying skill of decision-making in games to be reflected in video-based test performance. Unlike Nazarudin et al. (2015), there was no significant correlation for decision-making performance in each video-based test compared to that of in-game performance. Whereas Study 2 demonstrates 360° VR and match broadcast video-based tests have the sensitivity to differentiate between elite and amateur skill levels, Study 3 reports these tests in their current form do not transfer to match performance of elite officials. Given the theoretical notion that more representative tasks transfer to competition (Hadlow et al., 2018), 360° VR performance was hypothesised to have a strong relationship to on-field decision-making. Beyond officials, few studies in athletes assess transfer of decision-making performance in off-field tasks to on-field (Larkin et al., 2015; Renshaw et al., 2018), despite this commonly being noted as a requirement for future research (Paradis et al., 2016; Renshaw et al., 2018). Although past research has used high performance youth athletes to film scenarios for elite level soccer referees (Put et al., 2014), this may not be appropriate for Australian football umpires. This could be due to several factors such as the youth players in the video footage limiting the amount of task and domain-specific information for the elite umpires, which is key to differentiate between expert-based skill levels (Mann, Williams, Ward, & Janelle, 2007). The type of players adjudicated is a key constraint of competition, and may have limited the representativeness of this method. Match broadcast video, however, presents video specific to the elite population (i.e., elite AFL games), and has demonstrated transfer between video-based and in-game decision-making performance of elite Australian football players (Breed, Mills, & Spittle, 2018). No such relationship was observed in Study 3, suggesting match broadcast and 360° VR in their current forms are not transferrable to in-game performance. Although the 360° VR used small-sided games, the number of players in each decision-making scenario was designed to representative of the number of players in a game, and consistent with the player numbers for scenarios in the match broadcast video.

Finally, this thesis examined the use of both 360° VR and match broadcast in a video-based training program to develop decision-making skill. Few video-based training studies have established the reliability and validity of their testing tools before testing (Larkin et al., 2015), so performance changes in response to a training program in Study 4 were assessed using the reliable and valid video-based tests of Study 2. Although Study 3 suggests these tests do not have the sensitivity to differentiate elite performers, Study 2 identified both as reliable measures with the ability to differentiate between amateur and elite performers. It was hypothesised that 360° VR would be a superior training stimulus to match broadcast video due to the higher ecological validity observed in Study 2, and the levels of embodiment in virtual reality technologies leading to stronger, long-term behavioural changes (Kilteni et al., 2012). Similar to past research examining the use of 360° VR as a training stimulus (Pagé et al., 2019; Panchuk et al., 2018), the results of this study were promising for the use of this technology based on the difference between 360° VR and control groups and the higher ratings of game-likeness, enjoyment and relevance. There were no significant decision-making performance changes between 360° VR and match broadcast video groups. The 360° VR group scored statistically significantly higher than the control group in the retention test, with no other between-group differences observed. As this difference was only observed at the retention testing point, this may be due to the greater sense of embodiment in the 360° VR intervention group, which can increase the chances of long-term behavioural changes (Bohil et al., 2011). These long term behavioural changes such as the differences observed at the retention test may also indicate 360° VR is a more implicit type of learning, which requires longer practice periods to improve performance (Smeeton et al., 2015). This is consistent with studies suggesting the stereovision (i.e., binocular vision) afforded through a HMD leads to stronger levels of embodiment using this technology than screen-based modes such as match broadcast video (Faure et al., 2020; Kilteni et al., 2012). From a decision-making performance standpoint, it remains unclear whether 360° VR is a stronger perceptual-cognitive training tool than match broadcast video given no differences were observed in this study between groups. Study 4 highlights that participants considered 360° VR to be a more relevant and enjoyable training tool than match broadcast video. Given the high levels of fidelity and enjoyment of the participants using 360° VR, this may be a novel training method for players when training perceptual-cognitive skills in the absence of a movement

response. Similarly, Panchuk et al. (2018) reported positive feedback from the participants in a 360° VR training intervention for basketball decision-making. Training activities with high perceptions of relevance, enjoyment, concentration, and effort are important components of deliberate practice design (Catteeuw et al., 2009; MacMahon et al., 2007a). This suggests that, with further technological developments and understanding, 360° VR could be incorporated into program development for sports officials.

7.6. Practical implications

Results of this thesis suggest that both 360° VR and match broadcast video-based testing methods may be appropriate decision-making assessment tools for Australian football umpires, yet there are inherent limitations of using this video that practitioners must be aware of. Study 2 demonstrated the reliability of the two video modes over multiple testing occasions. In a practical setting, it is important to monitor performance changes over time, especially in situations where a participant may be injured and cannot be assessed in-game. Such scenarios warrant the use of video-based testing tools. In addition, 360° VR and match broadcast video may be used as talent identification tools given the strong construct validity evident in Study 2. As Study 1 highlights, match broadcast video-based tests can differentiate between amateur and elite officials in a variety of officiating populations. Findings of Study 2 outline 360° VR can similarly identify expert-based differences. For example, an official who is currently in an amateur performance group, but scoring at the level of an elite umpire on a valid video-based test, may be identified for a higher performance level. Although these video-based tests have the sensitivity to distinguish between skill levels, Study 3 suggests these video-based tests may not be able to distinguish between participants of an elite population, as assessed through in-game decision-making accuracy. Therefore, practitioners may consider using both video modes as a talent identification tool to prospectively identify elite officials, but must consider the limited sensitivity when analysing within-group differences of an elite population.

Study 4 aimed to investigate the effectiveness of 360° VR and match broadcast video in developing decision-making skill over a 5-week training intervention. Given the importance of decision-making to Australian football umpiring performance (Kittel et al., 2019b), it is imperative to identify novel methods to improve this skill. The findings of Study 4 suggest 360° VR is more beneficial than no video-based training (i.e., a control group). It must be noted this difference occurred at the retention testing point (four weeks post-intervention) in the 360° VR test, suggesting the difference between the 360° VR and control groups could be test-specific and are not immediately evident following the intervention. An important consideration for practitioners is the perceived greater relevance and enjoyment when using 360° VR compared to match broadcast video. Anecdotally, participants remarked on how the 360° VR was a more enjoyable tool with greater game-likeness (as assessed in the ecological validity/psychological fidelity scale) due to the first-person perspective and novel nature of this tool. The participants also did not consider this to require greater concentration or cognitive effort than match broadcast video. Although further research is required to determine the effectiveness of 360° VR, participants had greater satisfaction when using it compared to match broadcast video. Finally, rule changes between and within-season in sports such as Australian football require officials to apply these new interpretations in off-field settings before matches. Given Studies 2 and 4 report 360° VR to have higher game-likeness, this technology may be used as a training tool to assist officials in identifying different cues in newer rule interpretations.

7.7. Limitations of the thesis

There are several limitations of the current findings that need to be considered in future studies. Firstly, the participants captured in the 360° VR filming process were high performance youth players. Although these are the level amateur officials in Studies 2 & 4 would normally umpire, they are not the level normally officiated by the elite participants in Studies 2 & 3. This may explain why there was no significant difference in game-likeness rating (i.e., ecological validity) between 360° VR and match broadcast video for the elite participants of Study 2. Furthermore, it is hypothesised the athletes filmed

for the 360° VR (i.e., high performance junior umpires) contributed to the lack of significant findings observed in Study 3 comparing game and video-based decision-making performance. This may have limited the domain and context-specific information required to differentiate between participants (Mann et al., 2007). It would have been an ideal approach to film 360° VR video of elite Australian football players (i.e., from an AFL club). Recruitment of AFL clubs for filming was attempted, but not approved due to ethical/privacy concerns of filming closed training sessions and the potential disruption to their regimented training schedule. Therefore, high performance youth participants were recruited given similar approaches in past research (Put et al., 2014). The use of training video rather than match footage, may be considered a limitation for 360° VR despite this being used in previous studies using this technology (Pagé et al., 2019; Panchuk et al., 2018). Given representative design encapsulates the similarity of constraints in off-field and on-field settings (Pinder et al., 2011), a key constraint missing is the elite participants officiating scenarios of non-elite players. Current technological constraints limit the ability to capture in-game 360° VR, yet this may be considered as technology progresses as this would be a more representative approach. For each of the video modes, the final frame of the clip was presented for 0.5 sec following each clip. Although this was presented for a very small amount of time, research has shown that the final frame may provide additional information and is a consideration for future research using video-based training (Ryu, Abernethy, Mann & Poolton, 2015).

For the training intervention in Study 4, it is acknowledged that the overall length of the intervention may not have been long enough to elicit more significant changes. The five weeks was based on previous video-based training in Australian football players where five weeks was sufficient for positive training adaptations (Lorains et al., 2013a). As the scenarios were not scripted for the 360° VR library, unlike past research presenting tactical decisions for basketball players (Pagé et al., 2019; Panchuk et al., 2018), this limited the total number of videos available for the testing battery used in Studies 2, 3 and 4 and the training intervention battery of Study 4. By filming situations that were not scripted, this increased the fidelity and representativeness of the task by making the intensity more game-like and naturalistic. The training videos used in Study 4 could not be repeated in the testing battery. A number of factors limited the amount of 360° VR videos collected, including adequate

tackling scenarios where a decision could be made, clarity of decisions (e.g., distance from camera, light, body position of players etc.) and providing a distribution of different decisions that Australian football umpires make in games. There is a significant range of decisions possible for any tackling situation in Australian football (Larkin et al., 2018a), with the testing and training batteries covering each of the decisions possible in these scenarios.

Finally, it was not possible to assess the transfer of decision-making changes from the training intervention to on-field performance in Study 4. This is an important consideration of perceptual training tools (Hadlow et al., 2018), yet is not commonly quantified in training studies (Larkin et al., 2015; Renshaw et al., 2018). The assessment of transfer conducted in Study 3 used a rigorous in-game assessment procedure, where umpires were assessed over multiple games and decisions were reviewed by multiple coaches using quality match footage including additional camera angles not used in typical television broadcasts. This robust approach was not available to reliably measure in-game performance changes of the amateur officials in Study 4, as this group may not receive coaching/assessment each game and when they do, often there are different coaches who subjectively rate performance leading to inconsistencies. The games officiated by this group are not televised and therefore could not be scrutinised post-hoc as per the elite officials in Study 3. There has been suggestions more research should investigate elite participants' video-based training (Farrow et al., 2018). Most video-based training studies, however, do not include elite participants (Larkin et al., 2015), which is due to access to participants and the potential disruption the intervention may create leading to decreased game performance.

7.8. Future research directions

This thesis has advanced the literature surrounding video-based testing in sports officials by addressing several gaps identified in Study 1. A key premise of video-based testing is the isolation of decision-making skill for reliable measurement. Isolating decision-making skill may not be ideal as this presents a decontextualized assessment of this skill (O'Brien & Rynne, 2020). There are opportunities for video-based approaches to be more representative of the competitive environment for both testing and training purposes, by incorporating more of the constraints from competition (Pinder et al., 2015), which may then provide more contextualised approaches. There are, however, several other factors which may influence decision-making in-game and need to be managed by officials, such as position on the field (Corrigan et al., 2019); crowd noise (Balmer et al., 2007); physical exertion/fatigue (Bloß, Schorer, Loffing, & Büsch, 2020); managing interactions with players (Cunningham, Simmons, & Mascarenhas, 2018); contextual judgements (Burnett, Bishop, Ashford, Williams, & Kinrade, 2017); and sources of stress (Anshel, Kang, & Jubenville, 2013). Samuel, Galily, Guy, Sharoni, and Tenenbaum (2019) incorporated multiple constraints such as physical exertion and contextual judgement into a decision-making simulation. A more representative approach may be achieved by including multiple constraints such as those listed above into off-field testing and training protocols. A more representative approach would theoretically lead to subsequent performance improvements in competition, and performance in reliable off-field testing may more closely replicate in-game performance.

Training interventions such as Study 4 can be advanced in a number of ways. As highlighted, non-elite participants are commonly investigated in video-based training in general (Larkin et al., 2015), plus previous 360° VR interventions (Pagé et al., 2019; Panchuk et al., 2018). Future studies may attempt to recruit elite participants to determine whether there is a ceiling effect in this population, or video-based training can provide added training benefit. Video-based training is often short in duration (i.e., several weeks), yet further research could implement a longitudinal intervention such as over the course of a full year (Farrow et al., 2018). Such approaches may lead to stronger retention and transfer of training adaptations. Transfer of decision-making skill improvements post-intervention should also

be assessed, with participant groups recruited which have the capability to reliably measure on-field performance.

Results from this thesis suggest that 360° VR is a more ecologically valid tool than match broadcast video due to the stronger game-likeness ratings provided in Studies 2 and 4. It is interesting to note no differences in game-likeness were observed within the elite population in Study 2. Despite the 360° VR presenting non-game (i.e., training footage) of a lower performance group than they normally officiate, they did not consider the elite game footage of match broadcast video to be more game-like. Anecdotally, participants attributed their ratings to the first-person perspective being more similar to in-game and the ability to scan the environment with head movements. The head movements allowed for stronger behavioural correspondence which is a key consideration of representative tasks (Hadlow et al., 2018; Pinder et al., 2011). Further research could investigate why participants considered this perspective to be game-like, using methods such as concurrent verbalisation of perceptual-cognitive processes similar to Larkin et al. (2018a). Additionally, eye-tracking could be utilised to understand the differences of gaze behaviour between screen-based approaches such as match broadcast video in comparison to 360° VR presented through a HMD.

Advancements in technology may allow for different video methods to be examined for testing and training decision-making skills, particularly within officials. While this thesis examined 360° VR under the premise that it provides a more representative and ecologically valid decision-making tool, there may be stronger methods. For example, future studies could combine the strengths of first-person perspective in 360° VR with the elite game footage of match broadcast video, where elite officials wear a camera in-game. This would directly represent the perceptual information received in-game. This first-person game footage could be presented on a screen such as match broadcast video, or technological advancements may allow for 360° VR cameras to be worn on officials in competitive games. Presenting in-game 360° VR video would allow participants to freely scan the environment to identify decisions that may be out of the view of a standard camera. Such an approach would combine the ecological validity of first-person perspective with the behavioural correspondence of the head movements contributing to stronger representativeness.

7.9. Conclusion

In summary, this thesis provides a significant contribution to the literature in several ways. Officials play a crucial role in all sports yet remain significantly under-researched in comparison to players. In particular, their knowledge and application of the law (i.e., decision-making) is their most important skill given their decisions may influence the outcome of a match. This thesis presents greater understanding of how this skill is measured in a controlled off-field environment through video-based testing, with these studies summarised in the form of a systematic review (Study 1). Given video-based tests are commonly presented in the literature, this systematic review presents recommendations for future research in this area.

This thesis attempted to address the gaps identified in Study 1 through Studies 2 and 3. These include examination of an officiating group beyond soccer referees, establishing the reliability and validity, transfer to on-field performance, and use of other video modes than match broadcast footage. Video-based tests were developed based on the limitations of previous research. 360° VR and match broadcast both demonstrated strong construct validity and test re-test reliability. Study 2 was the first investigation to establish the reliability and validity of 360° VR as a decision-making assessment tool in sport. Results of Study 3, however, suggest neither video-based test (i.e., match broadcast or 360° VR) had the sensitivity to differentiate decision-making skill within an elite population. Study 2 identified 360° VR to be more game-like than match broadcast video, suggesting this method may be a more appropriate training tool for decision-making, particularly in amateur officials. Although research has suggested 360° VR is a suitable off-field decision-making tool, Study 2's game-likeness scale is the first to quantify this as a more ecologically valid tool for decision-making compared to match broadcast.

A significant contribution of this thesis is the use of 360° VR technology, particularly as a decision-making training tool. The results do not support the current 360° VR intervention as a statistically stronger training tool than match broadcast, however, there are several results which suggest 360° VR is a promising technology. The findings support the notion virtual reality such as 360° VR may provide strong immersion and presence, given the greater decision-making retention of the 360°

VR group in comparison to no training (i.e., control group). Therefore, 360° VR may assist with stronger long-term learning adaptations than match broadcast video, however no such differences between the two video modes were observed in Study 4. In particular, 360° VR appears to be practically useful for practitioners, given Study 4 identified this to be also a more enjoyable and relevant training tool than match broadcast. This finding has implications for deliberate practice approaches, where training should have high levels of enjoyment, relevance, concentration, and effort. Given 360° VR is an emerging technology, there are inherent limitations such as those described. Future studies could build on this work to present decision-making testing and training tools which are more ecologically valid and/or representative of the sporting environment. Although the participants were officials, the findings of this thesis may translate to the wider sporting community such as players.

16)

Chapter 8: References

Abernethy, B., & Russell, D. G. (1987). The relationship between expertise and visual search strategy in a racquet sport. *Human Movement Science, 6*(4), 283-319. https://doi.org/10.1016/0167-9457(87)90001-7

Abernethy, B., Thomas, K. T., & Thomas, J. T. (1993). Strategies for improving understanding of motor expertise [or mistakes we have made and things we have learned!!]. *Advances in Psychology, 102*, 317-356. https://doi.org/10.1016/S0166-4115(08)61478-8

Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-based Instruction, 15*, 40-47.

Anshel, M. H., Kang, M., & Jubenville, C. (2013). Sources of acute sport stress scale for sports officials: Rasch calibration. *Psychology of Sport and Exercise, 14*(3), 362-370. http://dx.doi.org/10.1016/j.psychsport.2012.12.003

Araújo, D., Davids, K., & Hristovski, R. (2006). The ecological dynamics of decision making in sport. *Psychology of Sport and Exercise, 7*(6), 653-676. https://doi.org/10.1016/j.psychsport.2006.07.002

Araujo, D., Davids, K., & Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: Comment on Rogers, Kadar, and Costall (2005). *Ecological Psychology, 19*(1), 69-78. https://doi.org/10.1080/10407410709336951

Baker, J., Cote, J., & Abernethy, B. (2003). Sport-specific practice and the development of expert decision-making in team ball sports. *Journal of Applied Sport Psychology, 15*(1), 12-25. https://doi.org/10.1080/10413200305400

Baker, J., Côté, J., & Abernethy, B. (2003). Learning from the experts: Practice activities of expert decision makers in sport. *Research Quarterly for Exercise and Sport, 74*(3), 342-347. https://doi.org/10.1080/02701367.2003.10609101

Balmer, N. J., Nevill, A. M., Lane, A. M., & Ward, P. (2007). Influence of crowd noise on soccer refereeing consistency in soccer. *Journal of Sport Behavior, 30*(2), 130.

Bar-Eli, M., Plessner, H., & Raab, M. (2011). *Judgment, decision-making and success in sport*: John Wiley & Sons.

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

Bideau, B., Kulpa, R., Vignais, N., Brault, S., Multon, F., & Craig, C. (2010). Using virtual reality to analyze sports performance. *IEEE Computer Graphics and Applications, 30*(2), 14-21. https://doi.org/10.1109/MCG.2009.134

Bird, J. M. (2020). The use of virtual reality head-mounted displays within applied sport psychology. *Journal of Sport Psychology in Action, 11*(2), 115-128. https://doi.org/10.1080/21520704.2018.1563573

Blair, M. R., Elsworthy, N., Rehrer, N. J., Button, C., & Gill, N. D. (2018). Physical and physiological demands of elite rugby union officials. *International Journal of Sports Physiology and Performance, 13*(9), 1199-1207. https://doi.org/10.1123/ijspp.2017-0849

Bloß, N., Schorer, J., Loffing, F., & Büsch, D. (2020). Physical load and referees' decision-making in sports games: A scoping review. *Journal of Sports Science and Medicine, 19*(1), 149.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience, 12*(12), 752-762. https://doi.org/10.1038/nrn3122

Bonney, N., Berry, J., Ball, K., & Larkin, P. (2019). Australian football skill-based assessments: A proposed model for future research. *Frontiers in Psychology, 10*, 429. https://doi.org/10.3389/fpsyg.2019.00429

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*(1), 93-99.

Brault, S., Bideau, B., Kulpa, R., & Craig, C. M. (2012). Detecting deception in movement: the case of the side-step in rugby. *PLoS One, 7*(6), e37494. https://doi.org/10.1371/journal.pone.0037494

Brault, S., Kulpa, R., Duliscouët, L., Marin, A., & Bideau, B. (2015). Virtual kicker vs. real goalkeeper in soccer: a way to explore goalkeeper's performance. *Movement Sport Sciences*(3), 79-88. https://doi.org/10.1051/sm/2015026

Breed, R., Mills, O., & Spittle, M. (2018). Can video-based perceptual-cognitive tests differentiate between skill level, player position, and experience in elite Australian football? *Journal of Expertise, 1*(1), 79-93.

Brightmore, A., O'Hara, J., Till, K., Cobley, S., Hubka, T., Emmonds, S., & Cooke, C. (2016). Movement and physiological demands of Australasian national rugby league referees. *International Journal of Sports Physiology and Performance, 11*(8), 1080-1087. https://doi.org/10.1123/ijspp.2015-0415

Broadbent, D. P., Causer, J., Williams, A. M., & Ford, P. R. (2015). Perceptual-cognitive skill training and its transfer to expert performance in the field: Future research directions. *European Journal of Sport Science, 15*(4), 322-331. https://doi.org/10.1080/17461391.2014.957727

Burnett, A. M., Bishop, D. T., Ashford, K. J., Williams, A. M., & Kinrade, N. P. (2017). Decision-making of English Netball Superleague umpires: Contextual and dispositional influences. *Psychology of Sport and Exercise, 31*, 52-60. http://dx.doi.org/10.1016/j.psychsport.2017.04.003

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

Catteeuw, P., Gilis, B., García-Aranda, J.-M., Tresaco, F., Wagemans, J., & Helsen, W. (2010a). Offside decision making in the 2002 and 2006 FIFA World Cups. *Journal of Sports Sciences, 28*(10), 1027-1032. https://doi.org/10.1080/02640414.2010.491084

Catteeuw, P., Gilis, B., Jaspers, A., Wagemans, J., & Helsen, W. (2010b). Training of perceptual-cognitive skills in offside decision making. *Journal of Sport and Exercise Psychology, 32*(6), 845-861. https://doi.org/10.1123/jsep.32.6.845

Catteeuw, P., Gilis, B., Wagemans, J., & Helsen, W. (2010c). Perceptual-cognitive skills in offside decision making: Expertise and training effects. *Journal of Sport and Exercise Psychology, 32*(6), 828-844. https://doi.org/10.1123/jsep.32.6.828

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

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98-101. https://doi.org/10.1111%2F1467-8721.ep10768783

Correia, V., Araújo, D., Cummins, A., & Craig, C. M. (2012). Perceiving and acting upon spaces in a VR rugby task: expertise effects in affordance detection and task achievement. *Journal of Sport and Exercise Psychology, 34*(3), 305-321. https://doi.org/10.1123/jsep.34.3.305

Corrigan, S. L., Dwyer, D. B., Harvey, B., & Gastin, P. B. (2019). The influence of match characteristics and experience on decision-making performance in AFL umpires. *Journal of Science and Medicine in Sport, 22*(1), 112-116. https://doi.org/10.1016/j.jsams.2018.06.005

Craig, C. (2013). Understanding perception and action in sport: how can virtual reality technology help? *Sports Technology, 6*(4), 161-169. https://doi.org/10.1080/19346182.2013.855224

Cunningham, I., Simmons, P., & Mascarenhas, D. (2018). Sport officials' strategies for managing interactions with players: Face-work on the front-stage. *Psychology of Sport and Exercise, 39*, 154-162. https://doi.org/10.1016/j.psychsport.2018.08.009

Cunningham, I., Simmons, P., Mascarenhas, D., & Redhead, S. (2014). Skilled interaction: concepts of communication and player management in the development of sport officials. *International Journal of Sport Communication, 7*(2), 166-187. https://doi.org/10.1123/IJSC.2013-0098

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26): Sage publications.

Dicks, M., Button, C., & Davids, K. (2010). Availability of advance visual information constrains association-football goalkeeping performance during penalty kicks. *Perception, 39*(8), 1111-1124. https://doi.org/10.1068%2Fp6442

Dicks, M., Davids, K., & Button, C. (2009). Representative task design for the study of perception and action in sport. *International Journal of Sport Psychology, 40*(4), 506-524.

Downward, P., & Jones, M. (2007). Effects of crowd size on referee decisions: Analysis of the FA Cup. *Journal of Sports Sciences, 25*(14), 1541-1545. https://doi.org/10.1080/02640410701275193

Düking, P., Holmberg, H.-C., & Sperlich, B. (2018). The potential usefulness of virtual reality systems for athletes: A short SWOT analysis. *Frontiers in Physiology, 9*, 128. https://doi.org/10.3389/fphys.2018.00128

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

Elsworthy, N., & Dascombe, B. J. (2011). The match demands of Australian rules football umpires in a state-based competition. *International Journal of Sports Physiology and Performance, 6*(4), 559-571. https://doi.org/10.1123/ijspp.6.4.559

Emmonds, S., O'Hara, J., Till, K., Jones, B., Brightmore, A., & Cooke, C. (2015). Physiological and movement demands of rugby league referees: Influence on penalty accuracy. *Journal of Strength and Conditioning Research, 29*(12), 3367-3374. https://doi.org/10.1519/JSC.0000000000001002

Erickson, G. B. (2007). *Sports vision: vision care for the enhancement of sports performance*: Elsevier Health Sciences.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Fadde, P. J., & Zaichkowsky, L. (2018). Training perceptual-cognitive skills in sports using technology. *Journal of Sport Psychology in Action, 9*(4), 239-248. https://doi.org/10.1080/21520704.2018.1509162

Farrow, D. (2007). The effect of viewing perspective on decision-making performance in water polo. *Journal of Sport and Exercise Psychology, 29*, S70-S71.

Farrow, D. (2013). Practice-enhancing technology: a review of perceptual training applications in sport. *Sports Technology, 6*(4), 170-176. https://doi.org/10.1080/19346182.2013.875031

Farrow, D., & Abernethy, B. (2002). Can anticipatory skills be learned through implicit video based perceptual training? *Journal of Sports Sciences, 20*(6), 471-485. https://doi.org/10.1080/02640410252925143

Farrow, D., Reid, M., Buszard, T., & Kovalchik, S. (2018). Charting the development of sport expertise: challenges and opportunities. *International Review of Sport and Exercise Psychology, 11*(1), 238-257. https://doi.org/10.1080/1750984X.2017.1290817

Farrow, D., & Robertson, S. (2016). Development of a skill acquisition periodisation framework for high-performance sport. *Sports Medicine, 47*(6), 1043-1054. https://doi.org/10.1007/s40279-016-0646-2

Faure, C., Limballe, A., Bideau, B., & Kulpa, R. (2020). Virtual reality to assess and train team ball sports performance: A scoping review. *Journal of Sports Sciences, 38*(2), 192-205. https://doi.org/10.1080/02640414.2019.1689807

Field, A. (2009). *Discovering statistics using SPSS*: Sage publications.

Freina, L., & Ott, M. (2015). *A literature review on immersive virtual reality in education: state of the art and perspectives.* Paper presented at the The International Scientific Conference eLearning and Software for Education.

Gabbett, T. J., Carius, J., & Mulvey, M. (2008). Does improved decision-making ability reduce the physiological demands of game-based activities in field sport athletes? *Journal of Strength and Conditioning Research, 22*(6), 2027-2035. https://doi.org/10.1519/JSC.0b013e3181887f34

Gadotti, I., Vieira, E., & Magee, D. (2006). Importance and clarification of measurement properties in rehabilitation. *Brazilian Journal of Physical Therapy, 10*(2), 137-146. https://doi.org/10.1590/S1413-35552006000200002

Ghasemi, A., Momeni, M., Jafarzadehpur, E., Rezaee, M., & Taheri, H. (2011). Visual skills involved in decision making by expert referees. *Perceptual and Motor Skills, 112*(1), 161-171. https://doi.org/10.2466%2F05.22.24.27.PMS.112.1.161-171

Giulianotti, R. (2012). *Football*: Wiley Online Library.

Gorman, A. D., & Farrow, D. (2009). Perceptual training using explicit and implicit instructional techniques: Does it benefit skilled performers? *International Journal of Sports Science and Coaching, 4*(2), 193-208. https://doi.org/10.1260%2F174795409788549526

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

Gray, R. (2017). Transfer of training from virtual to real baseball batting. *Frontiers in Psychology, 8*, 2183. https://doi.org/10.3389/fpsyg.2017.02183

Gulec, U., Yilmaz, M., Isler, V., O'Connor, R. V., & Clarke, P. M. (2019). A 3D virtual environment for training soccer referees. *Computer Standards and Interfaces, 64*, 1-10. https://doi.org/10.1016/j.csi.2018.11.004

Gurusamy, K. S., Aggarwal, R., Palanivelu, L., & Davidson, B. R. (2009). Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database of Systematic Reviews*(1). https://doi.org/10.1002/14651858.CD006575.pub2

Hadlow, S. M., Panchuk, D., Mann, D. L., Portus, M. R., & Abernethy, B. (2018). Modified perceptual training in sport: A new classification framework. *Journal of Science and Medicine in Sport, 21*(9), 950-958. https://doi.org/10.1016/j.jsams.2018.01.011

Hancock, D. J., Bennett, S., Roaten, H., Chapman, K., & Stanley, C. (2020). An analysis of literature on sport officiating research. *Research Quarterly for Exercise and Sport*. https://doi.org/10.1080/02701367.2020.1756198

Hancock, D. J., Rix-Lièvre, G., & Côté, J. (2015). Citation network analysis of research on sport officials: A lack of interconnectivity. *International Review of Sport and Exercise Psychology, 8*(1), 95-105. https://doi.org/10.1080/1750984X.2015.1022202

Hancock, D. J., & Ste-Marie, D. M. (2013). Gaze behaviors and decision making accuracy of higher- and lower-level ice hockey referees. *Psychology of Sport and Exercise, 14*(1), 66-71. 10.1016/j.psychsport.2012.08.002

Haycraft, J. A., Kovalchik, S., Pyne, D. B., & Robertson, S. (2017). Physical characteristics of players within the Australian Football League participation pathways: a systematic review. *Sports Medicine-open, 3*(1), 1-16. https://doi.org/10.1186/s40798-017-0109-9

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247. https://psycnet.apa.org/doi/10.1037/1040-3590.7.3.238

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

Hopkins, W. G. (2015). Spreadsheets for analysis of validity and reliability. *Sportscience, 19*, 36-44.

Hopwood, M. J., Mann, D. L., Farrow, D., & Nielsen, T. (2011). Does visual-perceptual training augment the fielding performance of skilled cricketers? *International Journal of Sports Science and Coaching, 6*(4), 523-536. https://doi.org/10.1260%2F1747-9541.6.4.523

Jackson, R. C. (2018). Stepovers and signal detection: Response sensitivity and bias in the differentiation of genuine and deceptive actions. *Frontiers in Psychology, 9*, 2043. https://doi.org/10.3389/fpsyg.2018.02043

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies, 23*(4), 1515-1529. https://doi.org/10.1007/s10639-017-9676-0

Jessiman, S. W., Harvey, B., Corrigan, S. L., & Gastin, P. B. (2019). Training and competition activity profiles of Australian football field umpires. *Journal of Strength and Conditioning Research*. https://doi.org/10.1519/jsc.0000000000002926

Johansen, B. T., & 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.

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

Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments, 21*(4), 373-387. https://doi.org/10.1162/PRES_a_00124

Kittel, A., Elsworthy, N., & Spittle, M. (2019a). Incorporating perceptual decision-making training into high-intensity interval training for Australian football umpires. *Journal of Sports Sciences, 37*(1), 29-35. https://doi.org/10.1080/02640414.2018.1480257

Kittel, A., Larkin, P., Elsworthy, N., Lindsay, R., & Spittle, M. (2020). Effectiveness of 360° virtual reality and match broadcast video to improve decision-making skill. *Science and Medicine in Football*. https://doi.org/10.1080/24733938.2020.1754449

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. (2019b). Identification of key performance characteristics of elite Australian football umpires. *International Journal of Sports Science and Coaching, 14*(4), 490-497. https://doi.org/10.1177%2F1747954119845071

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. (2019c). Using 360° virtual reality as a decision-making assessment tool in sport. *Journal of Science and Medicine in Sport, 22*(9), 1049-1053. https://doi.org/10.1016/j.jsams.2019.03.012

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

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. (2020). Transfer of 360° virtual reality and match broadcast video-based tests to on-field decision-making. *Science and Medicine in Football*. https://doi.org/10.1080/24733938.2020.1802506

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Krenn, B. (2014). The impact of uniform color on judging tackles in association football. *Psychology of Sport and Exercise, 15*(2), 222-225. https://doi.org/10.1016/j.psychsport.2013.11.007

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

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174. https://doi.org/10.2307/2529310

Larkin, P., Berry, J., Dawson, B., & Lay, B. (2011). Perceptual and decision-making skills of Australian football umpires. *International Journal of Performance Analysis in Sport, 11*(3), 427-437. https://doi.org/10.1080/24748668.2011.11868562

Larkin, P., Mesagno, C., Berry, J., & Spittle, M. (2014a). Development of a valid and reliable video-based decision-making test for Australian football umpires. *Journal of Science and Medicine in Sport, 17*(5), 552-555. https://doi.org/10.1016/j.jsams.2013.08.001

Larkin, P., Mesagno, C., Berry, J., & Spittle, M. (2018a). Exploration of the perceptual-cognitive processes that contribute to in-game decision-making of Australian football umpires. *International Journal of Sport and Exercise Psychology, 16*(2), 112-124. https://doi.org/10.1080/1612197X.2016.1167760

Larkin, P., Mesagno, C., Berry, J., Spittle, M., & Harvey, J. (2018b). Video-based training to improve perceptual-cognitive decision-making performance of Australian football umpires. *Journal of Sports Sciences, 36*(3), 239-246. https://doi.org/10.1080/02640414.2017.1298827

Larkin, P., Mesagno, C., Spittle, M., & Berry, J. (2015). An evaluation of video-based training programs for perceptual-cognitive skill development. A systematic review of current sport-based knowledge. *International Journal of Sport Psychology, 46*(6), 555-586.

Larkin, P., O'Brien, B., Mesagno, C., Berry, J., Harvey, J., & Spittle, M. (2014b). Assessment of decision-making performance and in-game physical exertion of Australian football umpires. *Journal of Sports Sciences, 32*(15), 1446-1453. https://doi.org/10.1080/02640414.2014.896998

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

Light, R. L., Harvey, S., & Mouchet, A. (2014). Improving 'at-action'decision-making in team sports through a holistic coaching approach. *Sport, Education and Society, 19*(3), 258-275. https://doi.org/10.1080/13573322.2012.665803

Lorains, M., Ball, K., & MacMahon, C. (2013a). An above real time training intervention for sport decision making. *Psychology of Sport and Exercise, 14*(5), 670-674. https://doi.org/10.1016/j.psychsport.2013.05.005

Lorains, M., Ball, K., & MacMahon, C. (2013b). Expertise differences in a video decision-making task: Speed influences on performance. *Psychology of Sport and Exercise, 14*(2), 293-297. https://doi.org/10.1016/j.psychsport.2012.11.004

MacMahon, C., Helsen, W., Starkes, J., & Weston, M. (2007a). Decision-making skills and deliberate practice in elite association football referees. *Journal of Sports Sciences, 25*(1), 65-78. https://doi.org/10.1080/02640410600718640

MacMahon, C., Mascarenhas, D., Plessner, H., Pizzera, A., Oudejans, R., & Raab, M. (2014). *Sports officials and officiating: Science and practice*. Oxon: Routledge.

MacMahon, C., & Mildenhall, B. (2012). A Practical Perspective on Decision Making Influences in Sports Officiating. *International Journal of Sports Science and Coaching, 7*(1), 153-166. https://doi.org/10.1260%2F1747-9541.7.1.153

MacMahon, C., Starkes, J., & Deakin, J. (2007b). Referee decision making in a video-based infraction detection task: Application and training considerations. *International Journal of Sports Science and Coaching, 2*(3), 257-265. https://doi.org/10.1260%2F174795407782233164

MacMahon, C., & Ste-Marie, D. (2002). Decision-making by experienced rugby referees: Use of perceptual information and episodic memory. *Perceptual and Motor Skills, 95*(2), 570-572. https://doi.org/10.2466%2Fpms.2002.95.2.570

Mallo, J., Frutos, P. G., 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

Mann, D., Farrow, D., Shuttleworth, R., & Hopwood, M. (2009). The influence of viewing perspective on decision-making and visual search behaviour in an invasive sport. *International Journal of Sport Psychology, 40*(4), 546-564.

Mann, D., Williams, M., Ward, P., & Janelle, C. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of Sport and Exercise Psychology, 29*(4), 457-478. https://doi.org/10.1123/jsep.29.4.457

Marteniuk, R. G. (1976). *Information processing in motor skills*. New York: Holt, Rinehart and Winston.

Mascarenhas, D., Button, C., Hara, D., & Dicks, M. (2009). Physical performance and decision making in association football referees: A naturalistic study. *The Open Sports Sciences Journal, 2*(1), 1-9. http://dx.doi.org/10.2174/1875399X00902010001

Mascarenhas, D., Collins, D., & Mortimer, P. (2005a). The accuracy, agreement and coherence of decision-making in rugby union officials. *Journal of Sport Behavior, 28*(3), 253-271.

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

Mascarenhas, D., Collins, D., Mortimer, P., & Morris, B. (2005c). Training accurate and coherent decision making in rugby union referees. *The Sport Psychologist, 19*(2), 131-147. https://doi.org/10.1123/tsp.19.2.131

Masters, R. S. (1992). Knowledge, knerves and know-how: The role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. *British Journal of Psychology, 83*(3), 343-358. https://doi.org/10.1111/j.2044-8295.1992.tb02446.x

Matamala-Gomez, M., Donegan, T., Bottiroli, S., Sandrini, G., Sanchez-Vives, M. V., & Tassorelli, C. (2019). Immersive virtual reality and virtual embodiment for pain relief. *Frontiers in Human Neuroscience, 13*, 279. https://doi.org/10.3389/fnhum.2019.00279

McMenemy, K., & Ferguson, R. S. (2007). *A hitchhiker's guide to virtual reality*: AK Peters/CRC Press.

Michalski, S. C., Szpak, A., Saredakis, D., Ross, T. J., Billinghurst, M., & Loetscher, T. (2019). Getting your game on: Using virtual reality to improve real table tennis skills. *PloS One, 14*(9), 1-14. https://doi.org/10.1371/journal.pone.0222351

Miles, H. C., Pop, S. R., Watt, S. J., Lawrence, G. P., & John, N. W. (2012). A review of virtual environments for training in ball sports. *Computers & Graphics, 36*(6), 714-726. http://dx.doi.org/10.1016/j.cag.2012.04.007

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Morris, G., & O'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

Nazarudin, M. N., Abdullah, M. R., Suppiah, P. K., Fauzee, M. S. O., Parnabas, V., & Abdullah, N. M. (2015). Decision making and performance of Malaysian rugby sevens referees. *Movement, Health and Exercise, 4*(1), 61-75. https://doi.org/10.15282/mohe.v4i0.45

Neil, R., Bayston, P., Hanton, S., & Wilson, K. (2013). The influence of stress and emotions on association football referees' decision-making. *Sport and Exercise Psychology Review, 9*(2), 22-41.

Neumann, D. L., Moffitt, R. L., Thomas, P. R., Loveday, K., Watling, D. P., Lombard, C. L., . . . Tremeer, M. A. (2018). A systematic review of the application of interactive virtual reality to sport. *Virtual Reality, 22*(3), 183-198. https://doi.org/10.1007/s10055-017-0320-5

Nevill, A. M., Balmer, N. J., & Williams, A. M. (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

Nevill, A. M., Hemingway, A., Greaves, R., Dallaway, A., & Devonport, T. J. (2017). Inconsistency of decision-making, the Achilles heel of referees. *Journal of Sports Sciences, 35*(22), 2257-2261. https://doi.org/10.1080/02640414.2016.1265143

Neville, T. J., Salmon, P. M., & Read, G. J. (2016). Analysis of In-Game Communication as an Indicator of Recognition Primed Decision Making in Elite Australian Rules Football Umpires. *Journal of Cognitive Engineering and Decision Making, 11*(1), 81-96. https://doi.org/10.1177%2F1555343416672557

Neville, T. J., Salmon, P. M., & Read, G. J. (2018). Radio Gaga? Intra-team communication of Australian Rules Football Umpires–Effect of radio communication on content, structure and frequency. *Ergonomics, 61*(2), 313-328. https://doi.org/10.1080/00140139.2017.1353140

O'Brien, K. A., & Rynne, S. B. (2020). Seen but not heard: missing the mark in conceptualizing high performance officiating. *Sport in Society*. https://doi.org/10.1080/17430437.2020.1732930

O'Connor, D., Larkin, P., & Williams, A. M. (2017). What learning environments help improve decision-making? *Physical Education and Sport Pedagogy, 22*(6), 647-660. https://doi.org/10.1080/17408989.2017.1294678

Pagé, C., Bernier, P.-M., & Trempe, M. (2019). Using video simulations and virtual reality to improve decision-making skills in basketball. *Journal of Sports Sciences, 37*(21), 2403-2410. https://doi.org/10.1080/02640414.2019.1638193

Page, K., & Page, L. (2010). Alone against the crowd: Individual differences in referees' ability to cope under pressure. *Journal of Economic Psychology, 31*(2), 192-199. https://doi.org/10.1016/j.joep.2009.08.007

Panchuk, D., Klusemann, M. J., & Hadlow, S. M. (2018). Exploring the effectiveness of immersive video for training decision-making capability in elite, youth basketball players. *Frontiers in Psychology, 9*, 1-9. https://doi.org/10.3389/fpsyg.2018.02315

Paradis, K., Larkin, P., & O'Connor, D. (2016). The effects of physical exertion on decision-making performance of Australian football umpires. *Journal of Sports Sciences, 34*(16), 1535-1541. https://doi.org/10.1080/02640414.2015.1122205

Petit, J.-P., & Ripoll, H. (2008). Scene perception and decision making in sport simulation: a masked priming investigation. *International Journal of Sport Psychology, 39*(1), 1-19.

Petri, K., Ohl, C., Danneberg, M., Emmermacher, P., Masik, S., & Witte, K. (2018). Towards the usage of virtual reality for training in sports. *Biomedical Journal of Scientific and Technical Research, 7*(1), 1-3. https://doi.org/10.26717/BJSTR.2018.07.001453

Pinder, R. A., Davids, K. W., Renshaw, I., & Araújo, D. (2011). Representative learning design and functionality of research and practice in sport. *Journal of Sport and Exercise Psychology, 33*(1), 146-155. https://doi.org/10.1123/jsep.33.1.146

Pinder, R. A., Headrick, J., & Oudejans, R. R. (2015). Issues and challenges in developing representative tasks in sport. In *The Routledge Handbook of Sports Expertise* (pp. 269-281): Routledge.

Plessner, H., & Betsch, T. (2001). Sequential effects in important referee decisions: The case of penalties in soccer. *Journal of Sport and Exercise Psychology, 23*(3), 254-259. https://doi.org/10.1123/jsep.23.3.254

Poolton, J., Siu, C. M., & Masters, R. (2011). The home team advantage gives football referees something to ruminate about. *International Journal of Sports Science and Coaching, 6*(4), 545-552. https://doi.org/10.1260%2F1747-9541.6.4.545

Put, K., Baldo, M. V. C., Cravo, A. M., Wagemans, J., & Helsen, W. F. (2013a). Experts in offside decision making learn to compensate for their illusory perceptions. *Journal of Sport and Exercise Psychology, 35*(6), 576-584. https://doi.org/10.1123/jsep.35.6.576

Put, K., Wagemans, J., Jaspers, A., & Helsen, W. F. (2013b). Web-based training improves on-field offside decision-making performance. *Psychology of Sport and Exercise, 14*(4), 577-585. http://dx.doi.org/10.1016/j.psychsport.2013.03.005

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

Put, K., Wagemans, J., Spitz, J., Armenteros Gallardo, M., Williams, A. M., & Helsen, W. F. (2014). The use of 2D and 3D information in a perceptual-cognitive judgement task. *Journal of Sports Sciences, 32*(18), 1688-1697. https://doi.org/10.1080/02640414.2014.912760

Put, K., Wagemans, J., Spitz, J., Williams, A. M., & Helsen, W. F. (2015). Using web-based training to enhance perceptual-cognitive skills in complex dynamic offside events. *Journal of Sports Sciences, 34*(2), 181-189. https://doi.org/10.1080/02640414.2015.1045926

Raab, M. (2003). Implicit and explicit learning of decision making in sports is effected by complexity of situation. *International Journal of Sport Psychology*.

Raab, M., MacMahon, C., Avugos, S., & Bar-Eli, M. (2020). The referee's challenge: A threshold process model for decision making in sport games. *International Review of Sport and Exercise Psychology*, 1-21. https://doi.org/10.1080/1750984X.2020.1783696

Reger, G. M., Holloway, K. M., Candy, C., Rothbaum, B. O., Difede, J., Rizzo, A. A., & Gahm, G. A. (2011). Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. *Journal of Traumatic Stress, 24*(1), 93-96. https://doi.org/10.1002/jts.20574

Renden, P. G., Kerstens, S., Oudejans, R. R., & Cañal-Bruland, R. (2014). Foul or dive? Motor contributions to judging ambiguous foul situations in football. *European Journal of Sport Science, 14*(sup1), S221-S227. https://doi.org/10.1080/17461391.2012.683813

Renshaw, I., Davids, K., Araújo, D., Lucas, A., Roberts, W. M., Newcombe, D. J., & Franks, B. (2018). Evaluating weaknesses of "perceptual-cognitive training" and "brain training" methods in sport: An ecological dynamics critique. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.02468

Riva, G., Wiederhold, B. K., & Mantovani, F. (2019). Neuroscience of virtual reality: From virtual exposure to embodied medicine. *Cyberpsychology, Behavior, and Social Networking, 22*(1), 82-96. https://doi.org/10.1089/cyber.2017.29099.gri

Roca, A., & Williams, A. M. (2016). Expertise and the interaction between different perceptual-cognitive skills: Implications for testing and training. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00792

Russell, S., Renshaw, I., & Davids, K. (2019). How interacting constraints shape emergent decision-making of national-level football referees. *Qualitative Research in Sport, Exercise and Health, 11*(4), 573-588. https://doi.org/10.1080/2159676X.2018.1493525

Ryu, D., Abernethy, B., Mann, D. L., & Poolton, J. M. (2015). The contributions of central and peripheral vision to expertise in basketball: How blur helps to provide a clearer picture. *Journal of Experimental Psychology: Human Perception and Performance, 41*(1), 167. https://doi.org/10.1037/a0038306

Samuel, R. D., Galily, Y., Guy, O., Sharoni, E., & Tenenbaum, G. (2019). A decision-making simulator for soccer referees. *International Journal of Sports Science and Coaching, 14*(4), 480-489. https://doi.org/10.1177%2F1747954119858696

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

Shafizadeh, M., & Platt, G. K. (2012). Effect of verbal cueing on trajectory anticipation in the penalty kick among novice football goalkeepers. *Perceptual and Motor Skills, 114*(1), 174-184. https://doi.org/10.2466%2F05.23.25.PMS.114.1.174-184

Slack, L. A., Maynard, I. W., Butt, J., & Olusoga, P. (2013). Factors underpinning football officiating excellence: perceptions of English Premier League referees. *Journal of Applied Sport Psychology, 25*(3), 298-315. https://doi.org/10.1080/10413200.2012.726935

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology, 109*(3), 431-433. https://doi.org/10.1111/bjop.12305

Smeeton, N. J., Williams, A. M., Hodges, N. J., & Ward, P. (2005). The relative effectiveness of various instructional approaches in developing anticipation skill. *Journal of Experimental Psychology: Applied, 11*(2), 98. https://psycnet.apa.org/doi/10.1037/1076-898X.11.2.98

Souchon, N., Coulomb-Cabagno, G., Traclet, A., & Rascle, O. (2004). Referees' decision making in handball and transgressive behaviors: Influence of stereotypes about gender of players? *Sex Roles, 51*(7-8), 445-453. https://doi.org/10.1023/B:SERS.0000049233.28353.f0

Souchon, N., Livingstone, A. G., & Maio, G. R. (2013). The influence of referees' expertise, gender, motivation, and time constraints on decisional bias against women. *Journal of Sport and Exercise Psychology, 35*(6), 585-599. https://doi.org/10.1123/jsep.35.6.585

Spitz, J., Moors, P., Wagemans, J., & Helsen, W. F. (2018). The impact of video speed on the decision-making process of sports officials. *Cognitive Research: Principles and Implications, 3*(1), 1-10. https://doi.org/10.1186/s41235-018-0105-8

Spitz, J., Put, K., Wagemans, J., Williams, A. M., & Helsen, W. F. (2016). Visual search behaviors of association football referees during assessment of foul play situations. *Cognitive Research: Principles and Implications, 1*(1), 1-11. https://doi.org/10.1186/s41235-016-0013-8

Spitz, J., Put, K., Wagemans, J., Williams, A. M., & Helsen, W. F. (2017). Does slow motion impact on the perception of foul play in football? *European Journal of Sport Science, 17*(6), 748-756. https://doi.org/10.1080/17461391.2017.1304580

Spitz, J., Put, K., Wagemans, J., Williams, A. M., & Helsen, W. F. (2018). The role of domain-generic and domain-specific perceptual-cognitive skills in association football referees. *Psychology of Sport and Exercise, 34*(Supplement C), 47-56. https://doi.org/10.1016/j.psychsport.2017.09.010

Starkes, J. L., Deakin, J. M., Allard, F., Hodges, N. J., & Hayes, A. (1996). Deliberate practice in sports: What is it anyway? In K. A. Ericsson (Ed.), *The Road to Excellence: The Acquisition of Expert Performance in the Arts, Sciences, Sports and Games* (pp. 81-106). Mahwah, NJ: Lawrence Erlbaum.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, 42*(4), 73-93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Stoffregen, T. A., Bardy, B. G., Smart, L., & Pagulayan, R. (2003). On the nature and evaluation of fidelity in virtual environments. In L. Hettinger & M. Haas (Eds.), *Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues* (pp. 111-128). Mahwah, New Jersey.

Swann, C., Moran, A., & Piggott, D. (2015). Defining elite athletes: Issues in the study of expert performance in sport psychology. *Psychology of Sport and Exercise, 16, Part 1*, 3-14. http://dx.doi.org/10.1016/j.psychsport.2014.07.004

Thomas, J. R., Silverman, S., & Nelson, J. (2015). *Research methods in physical activity, 7E*: Human Kinetics.

van Biemen, T., Koedijker, J., Renden, P. G., & Mann, D. L. (2018). The effect of blurred perceptual training on the decision making of skilled football referees. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01803

Vickers, J. N. (2007). *Perception, cognition, and decision training: The quiet eye in action*: Human Kinetics.

Vignais, N., Kulpa, R., Brault, S., Presse, D., & Bideau, B. (2015). Which technology to investigate visual perception in sport: Video vs. virtual reality. *Human Movement Science, 39*, 12-26. https://doi.org/10.1016/j.humov.2014.10.006

Wagner-Egger, P., Gygax, P., & Ribordy, F. (2012). Racism in soccer? Perception of challenges of black and white players by white referees, soccer players, and fans. *Perceptual and Motor Skills, 114*(1), 275-289. https://doi.org/10.2466%2F05.07.17.PMS.114.1.275-289

Weston, M., Castagna, C., Impellizzeri, F. M., Bizzini, M., Williams, A. M., & 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, A. M., Davids, K., & Williams, J. G. P. (1999). *Visual perception and action in sport*: Taylor & Francis.

Williams, A. 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

Williams, A. M., & Ford, P. R. (2008). Expertise and expert performance in sport. *International Review of Sport and Exercise Psychology, 1*(1), 4-18. https://doi.org/10.1080/17509840701836867

Williams, A. M., & Ford, P. R. (2013). Game intelligence': Anticipation and decision making. In *Science and Soccer: Developing Elite Performers* (3 ed., pp. 105-121). Oxon: Routledge.

Williams, A. M., Ward, P., & Chapman, C. (2003). Training perceptual skill in field hockey: Is there transfer from the laboratory to the field? *Research Quarterly for Exercise and Sport, 74*(1), 98-103. https://doi.org/10.1080/02701367.2003.10609068

Williams, A. M., Ward, P., Smeeton, N., & Allen, D. (2004). Developing anticipation skills in tennis using on-court instruction: Perception versus perception and action. *Journal of Applied Sport Psychology, 16*(4), 350-360. https://doi.org/10.1080/10413200490518002

Wilson, A. W., & Mock, S. E. (2013). Association of certification level and assertiveness with accuracy of calls among ice hockey referees. *International Journal of Sports Science and Coaching*, *8*(3), 505-512. https://doi.org/10.1260%2F1747-9541.8.3.505

Woods, C. T., Raynor, A. J., Bruce, L., & McDonald, Z. (2016). Discriminating talent-identified junior Australian football players using a video decision-making task. *Journal of Sports Sciences*, *34*(4), 342-347. https://doi.org/10.1080/02640414.2015.1053512

Chapter 9: Appendices

Appendix A: Information statement (Capture of 360° VR)

INFORMATION TO PARTICIPANTS INVOLVED IN RESEARCH

You are invited to participate

You are invited to participate in a research project entitled *"Capturing 360° video footage of Australian Football"*.

This project is being conducted by a student researcher Mr Aden Kittel as part of a PhD study at Victoria University under the supervision of Assoc Prof. Michael Spittle and Dr. Paul Larkin from the College of Sport and Exercise Science Victoria University.

Project explanation

This project aims to capture 360° footage of Australian Football training drills and activities. Given the rise of virtual reality technology in recent years, 360° footage has been proposed to be a more life-like method for filming, compared to traditional 2-Dimensional techniques. This project will record 360° footage of Australian football training drills and activities that can be used in a later study investigating decision-making skill of Australian football umpires. This is the first research to examine the ability to capture 360° camera footage of team sport activity, including Australian football.

What will I be asked to do?

This research involves obtaining footage of training drills and activities of Australian football players using a 360° camera (360fly 4k Camera, 360fly, Canonsburg, USA).

A range of common training drills and activities (e.g., small sided games) as prescribed by the team's coaching staff, will be recorded using a 360° camera. The researchers will have no influence on the type of activities designed and conducted by the coaches. As training sessions are currently filmed from an aerial perspective in the grandstand, this project will look to situate a 360° camera in areas of the ground that replicate the position of an umpire in a game. This process will pose no additional burden to the players.

What will I gain from participating?

By participating in this study, this footage will be made available to you and your coaches for training purposes. Although training footage is used as part of standard practice, these videos will provide a new viewpoint of watching training drills and activities. This type of filming has not been used previously in sport research, and therefore you will have access to cutting edge sport technology.

How will the information I give be used?

This video footage will be used in a later research project exploring the assessment and development of decision-making skill in Australian football umpires. As such, this footage will only be used for the purpose of a testing and training tool for Australian football umpires at elite, sub-elite, and amateur levels. This footage will not be released to the public.

What are the potential risks of participating in this project?

There are no additional risks for participating in this project other than those present for normal training. As footage being recorded is standard training drills and activities, there will be no additional burden for participants.

How will this project be conducted?

This research will involve filming standard training drills and activities using a 360° camera approximately 15m away. To ensure an adequate amount of footage is captured for the follow-up research project, it is anticipated the filming process will be completed within 2-3 months (i.e., 1 session filmed per week for 8-12 weeks).

Who is conducting the study?

Assoc. Prof. Michael Spittle
Academic
Victoria University
Ph: 03 9919 9512
Email: Michael.Spittle@vu.edu.au

Mr Aden Kittel
PhD Candidate
Victoria University
Ph: 0402 636 419
Email: Aden.Kittel@live.vu.edu.au

Dr. Paul Larkin
Research Associate
Victoria University
Email. Paul.Larkin@vu.edu.au

Any queries about your participation in this project may be directed to the Chief Investigator listed above.
If you have any queries or complaints about the way you have been treated, you may contact the Ethics Secretary, Victoria University Human Research Ethics Committee, Office for Research, Victoria University, PO Box 14428, Melbourne, VIC, 8001, email researchethics@vu.edu.au or phone (03) 9919 4781 or 4461.

Appendix B: Consent form (Capture of 360° VR)

CONSENT FORM FOR PARTICIPANTS INVOLVED IN RESEARCH

INFORMATION TO PARTICIPANTS:
We would like to invite you to be a part of a study into developing *"Virtual Reality as a novel video-based method for decision-making in Australian football umpires"*.

This research aims to examine the efficacy of developing virtual reality footage obtained through 360° filming of Australian football drills. Virtual reality footage will be created using a 360° camera approximately 10m from training drills, which will be used for assessing and developing decision-making in Australian football umpires in a separate study. As this study involves only filming drills being conducted as part of scheduled training, there are no additional risks present for those participating.

CERTIFICATION BY PARTICIPANT

I, _____ (full name)

of _____ (suburb)

certify that I am voluntarily giving my consent to participate in the study.

CERTIFICATION BY PARENT/GUARDIAN

I, _____ (full name)

of _____ (suburb)

certify that I am giving my consent for my child to participate in this study being conducted at Victoria University by Associate Professor Michael Spittle, Dr Paul Larkin, Mr Aden Kittel.

I certify that the objectives of the study, together with any risks and safeguards associated with the procedures listed hereunder to be carried out in the research, have been fully explained to me by Mr Aden Kittel and that I freely consent to participation involving the below mentioned procedures:

- ☐ Filming of training drills (please tick)
- ☐ The use of this footage for decision-making assessment and training in Australian football umpires (please tick)

I certify that I have had the opportunity to have any questions answered and that I understand that I can withdraw from this study at any time and that this withdrawal will not jeopardise me in any way.

I have been informed that the information I provide will be kept confidential.

Signed: _____ (Participant) Signed: _____ (Parent; if under 18)

Date: _____ Date: _____

Any queries about your participation in this project may be directed to the researcher

Associate Professor Michael Spittle Ph: 03 9919 9512

If you have any queries or complaints about the way you have been treated, you may contact the Ethics Secretary, Victoria University Human Research Ethics Committee, Office for Research, Victoria University, PO Box 14428, Melbourne, VIC, 8001, email Researchethics@vu.edu.au or phone (03) 9919 4781 or 4461.

Appendix C: Information statement (Study 2)

INFORMATION TO PARTICIPANTS INVOLVED IN RESEARCH

You are invited to participate

You are invited to participate in a research project entitled "*Validity and reliability of video-based footage in Australian football umpires*".

This project is being conducted by a student researcher Mr Aden Kittel as part of a PhD study at Victoria University under the supervision of Assoc Prof. Michael Spittle and Dr. Paul Larkin from the College of Sport and Exercise Science Victoria University, and Dr Nathan Elsworthy from Central Queensland University.

Project explanation

Decision-making accuracy is considered the most important performance component for Australian football (AF) umpires, given the substantial impact each decision can have on the final outcome of a match. Although officiating actual match play is currently regarded as the best form of practice for umpires, there are only a finite number of games available to officiate. Therefore, it is imperative to investigate other possible testing procedures for umpire decision-making skill development. To replicate game-like performance, video-based methods are believed to promote the development of decision-making in officials. With the on-going development of video technology, it is also possible that 360° video virtual reality may be an appropriate method to assess decision-making skills. This research will develop a video-based (i.e., match broadcast footage) and a virtual reality (i.e., 360° video virtual reality) decision-making assessment tool. The overall aim of the project is to develop a valid and reliable decision-making testing tool for use in an off field, controlled environment.

What will I be asked to do?

Involves the completion of two video-based activities separated by two weeks. Each activity will involve watching Australian football footage through a virtual reality headset (i.e., 360° virtual reality), and on a computer screen (i.e., match broadcast footage), where you will be asked to provide a verbal free kick decision as would in a game. The duration of each testing session (two in total) will be approximately 45 min, and will be conducted at your regular training venue. Included within this 45 min test will be a 5 min familiarisation (i.e. presentation of 3 videos of each mode that are not used in the test itself) of each video mode prior to the test itself. This will allow you to be accustomed to the type of footage presented to you. Halfway through the testing session (i.e. between presentations of the two video modes) will be a 5 min break. The two testing sessions will be separated by two weeks. The virtual reality headset is designed to accommodate individuals with glasses also.

What will I gain from participating?

It is anticipated that this project will have several benefits to you. Firstly, you will be able to experience virtual reality as a novel mode for Australian football decision-making. This research will evaluate the appropriateness of different video-based methods in assessing decision-making skill. This knowledge will assist in game decision-making of Australian football umpires to improve overall performance.

How will the information I give be used?

Results from this study will be used to enhance current Australian football umpires testing practices. This knowledge will enhance talent identification, and selection processes of Australian football umpires.

Whilst results will be published in scientific journals, personal information will not be disclosed and all data will be coded to prevent identification of specific individuals. Specifically, only group-based averages will be used in any publications and reports emanating from this project. No individual data will be reported.

What are the potential risks of participating in this project?

The associated risks of this project are minimal as the testing and training programs will be conducted at your regular training venue. Virtual reality does occasionally cause motion sickness, however, you will remain seated which should negate any symptoms. If you do not wish to participate in this study there will be no ramifications in terms of the level of service, or have any influence on current or future associations with either Victoria University or the AFL umpiring department.

How will this project be conducted?

This study will involve completing two tests with each test using two different conditions (360° virtual reality and match broadcast footage). Each video mode (i.e. 360° virtual reality and match broadcast footage) will contain 60 clips per test, where you will be asked to provide a decision on what your decision is (e.g. "play on", "holding the ball" etc.). Each test (which incorporates both video modes) will take approximately 45 min. The tests will be separated by two weeks. This study will be administered by the student researcher (Mr Aden Kittel).

Who is conducting the study?

Assoc. Prof. Michael Spittle
Academic
Victoria University
Ph: 03 9919 9512
Email: Michael.Spittle@vu.edu.au

Mr Aden Kittel
PhD Candidate
Victoria University
Ph: 0402 636 419
Email: Aden.Kittel@live.vu.edu.au

Dr. Paul Larkin
Research Associate
Victoria University
Email. Paul.Larkin@vu.edu.au

Dr Nathan Elsworthy
Lecturer
CQUniversity
Ph: 07 4940 3382
Email: N.Elsworthy@cqu.edu.au

Any queries about your participation in this project may be directed to the Chief Investigator listed above.
If you have any queries or complaints about the way you have been treated, you may contact the Ethics Secretary, Victoria University Human Research Ethics Committee, Office for Research, Victoria University, PO Box 14428, Melbourne, VIC, 8001, email researchethics@vu.edu.au or phone (03) 9919 4781 or 4461.

Appendix D: Consent form (Study 2)

CONSENT FORM FOR PARTICIPANTS INVOLVED IN RESEARCH

INFORMATION TO PARTICIPANTS:
We would like to invite you to be a part of a study into *"Validity and reliability of video-based footage in Australian football umpires".*

This study will involve completing two tests with each test using two different conditions (360° virtual reality and match broadcast footage). Each video mode (i.e. 360° virtual reality and match broadcast footage) will contain 60 clips per test, where you will be asked to provide a decision on what your decision is (e.g. "play on", "holding the ball" etc.). Each test (which incorporates both video modes) will take approximately 45 min. The tests will be separated by two weeks. This study will be administered by the student researcher (Mr Aden Kittel).There are minimal risks associated with this research. While unlikely, you may feel symptoms of motion sickness while using virtual reality.

CERTIFICATION BY PARTICIPANT

I, _____(full name)

of _____(suburb)

certify that I am at least 18 years old and that I am voluntarily giving my consent to participate in the study:
Validity and reliability of video-based footage in Australian football umpires
being conducted at Victoria University by Associate Professor Michael Spittle, Dr Paul Larkin, Mr Aden Kittel, and Dr Nathan Elsworthy.

I certify that the objectives of the study, together with any risks and safeguards associated with the procedures listed hereunder to be carried out in the research, have been fully explained to me by Mr Aden Kittel, and that I freely consent to participation involving the below mentioned procedures:

☐ Participating in video-based testing (please tick).

I certify that I have had the opportunity to have any questions answered and that I understand that I can withdraw from this study at any time and that this withdrawal will not jeopardise me in any way.

I have been informed that the information I provide will be kept confidential.

Signed: _____a

Date: _____a

Any queries about your participation in this project may be directed to the researcher
Associate Professor Michael Spittle
Ph: 03 9919 9512

If you have any queries or complaints about the way you have been treated, you may contact the Ethics Secretary, Victoria University Human Research Ethics Committee, Office for Research, Victoria University, PO Box 14428, Melbourne, VIC, 8001, email Researchethics@vu.edu.au or phone (03) 9919 4781 or 4461.

Appendix E: Information statement (Study 3)

INFORMATION TO PARTICIPANTS INVOLVED IN RESEARCH

You are invited to participate

You are invited to participate in a research project entitled "*Decision-making training for Australian football umpires*".

This project is being conducted by a student researcher Mr Aden Kittel as part of a PhD study at Victoria University under the supervision of Assoc Prof. Michael Spittle, Dr. Paul Larkin and Prof. Damian Farrow from the College of Sport and Exercise Science Victoria University, and Dr Nathan Elsworthy from Central Queensland University.

Project explanation

Decision-making accuracy is considered the most important performance component for Australian football (AF) umpires, given the substantial impact each decision can have on the final outcome of a match. Although officiating actual match play is currently regarded as the best form of practice for umpires, there are only a finite number of games available to officiate. Therefore, it is imperative to investigate other possible training methods for umpire decision-making skill development.
To replicate game-like performance, video-based methods (i.e., watching match vision) are believed to promote the development of decision-making in officials. With the on-going development of video technology, it is also possible that 360° video virtual reality may be a more effective method to develop decision-making skills. This research will develop a video-based (i.e., first-person match footage) and a virtual reality (i.e., 360° video virtual reality) decision-making training program. The overall aim of the project is to develop a training tool to improve match performance of Australian football umpires.

What will I be asked to do?

This study will include four decision-making tests (30 min each), and short training sessions (10 min each) delivered during training hours for two six-week blocks. The testing sessions will require you to watch individual clips of two different video methods (i.e., 360° video virtual reality and match broadcast footage), and provide a decision as you would in a game (e.g. "play on", "holding the ball" etc.). Each test will include a 5 min familiarisation including several clips to demonstrate the type of scenarios you will be watching. Following the first test, you will be provided 360° video virtual reality footage during your standard training hours as a coaching tool. These training sessions will be once per week (10 min each) for six consecutive weeks. You will complete another six week block four weeks later.

The process is as follows:

- Round 1: Pre-test (360° virtual reality and match broadcast test) (30 min)
- Rounds 2-7: Training twice per week (360° virtual reality) (10 min each)
- Round 8: Mid-test (360° virtual reality and match broadcast test) (30 min)
- Rounds 9-10: No video
- Rounds 11-16: Training twice per week (first-person match footage) (10 min each)
- Round 17: Post-test (360° virtual reality and match broadcast test) (30 min)
- Round 21: Retention-test (360° virtual reality and match broadcast test) (30 min)

Your match decision-making accuracy will be compared to the video-based test scores, to assess the transferability of the decision-making training to on-field performance. Your match decision-making and video-based test performance will be reported only as group averages, and will therefore not individually identify you.

What will I gain from participating?

It is anticipated that this project will have several benefits to you. Firstly, you will be able to experience virtual reality as a novel mode for Australian football decision-making. This research will evaluate the effectiveness of different video-based methods in developing decision-making skill. This knowledge will assist in game decision-making of Australian football umpires to improve overall performance.

How will the information I give be used?

Results from this study will be used to enhance current Australian football umpires training practices. This knowledge will enhance match performance, training, and talent identification of Australian football umpires.

No individual data will be reported. Whilst results will be published in scientific journals, personal information will not be disclosed and all data will be coded to prevent identification of specific individuals. Specifically, only group-based averages will be used in any publications and reports emanating from this project.

What are the potential risks of participating in this project?

The associated risks of this project are minimal as the testing and training programs will be conducted at your regular training venue. Virtual reality does occasionally cause motion sickness, however, you will remain seated which should negate any symptoms. If you do not wish to participate in this study there will be no ramifications in terms of the level of service, or have any influence on current or future associations with either Victoria University or the AFL umpiring department.

How will this project be conducted?

The structure of this project is as follows:

- Round 1: Pre-test (360° virtual reality and match broadcast test) (30 min)
- Rounds 2-7: Training twice per week (360° virtual reality) (10 min each)
- Round 8: Mid-test (360° virtual reality and match broadcast test) (30 min)
- Rounds 9-10: No video
- Rounds 11-16: Training twice per week (first-person match footage) (10 min each)
- Round 17: Post-test (360° virtual reality and match broadcast test) (30 min)
- Round 21: Retention-test (360° virtual reality and match broadcast test) (30 min)

This study will be administered by the student researcher (Mr Aden Kittel).

Who is conducting the study?

Assoc. Prof. Michael Spittle
Academic
Victoria University
Ph: 03 9919 9512
Email: Michael.Spittle@vu.edu.au

Mr Aden Kittel
PhD Candidate
Victoria University
Ph: 0402 636 419
Email: Aden.Kittel@live.vu.edu.au

Dr. Paul Larkin
Research Associate
Victoria University
Email. Paul.Larkin@vu.edu.au

Dr Nathan Elsworthy
Lecturer
CQUniversity
Ph: 07 4940 3382
Email: N.Elsworthy@cqu.edu.au

Professor Damian Farrow
Academic
Victoria University
Email. Damian.Farrow@vu.edu.au

Any queries about your participation in this project may be directed to the Chief Investigator listed above. If you have any queries or complaints about the way you have been treated, you may contact the Ethics Secretary, Victoria University Human Research Ethics Committee, Office for Research, Victoria University, PO Box 14428, Melbourne, VIC, 8001, email researchethics@vu.edu.au or phone (03) 9919 4781 or 4461.

Appendix F: Consent form (Study 3)

CONSENT FORM FOR PARTICIPANTS INVOLVED IN RESEARCH

INFORMATION TO PARTICIPANTS:

We would like to invite you to be a part of a study into *"Decision-making training for Australian football umpires".*

This study will involve completing decision-making testing (30 min each session) before and after a five week training program. Each video test will include two video modes (i.e. 360° virtual reality and match broadcast footage) containing 50 clips in total, where you will be asked to provide a decision on what your decision is (e.g. "play on", "holding the ball" etc.). A five week training program will immediately follow the first test, where you will be provided with decision-making scenarios once per week (10 min per session), for five weeks. This program will be followed by two more testing sessions as per the first test. The two final testing sessions will be one week, and four weeks following completion of the training program, respectively.

This study will be administered by the student researcher (Mr Aden Kittel).There are minimal risks associated with this research. While unlikely, you may feel symptoms of motion sickness while using virtual reality.

CERTIFICATION BY PARTICIPANT

I, _____ (full name)

of _____ (suburb)

certify that I am at least 18 years old and that I am voluntarily giving my consent to participate in the study:
Decision-making training for Australian football umpires
being conducted at Victoria University by Associate Professor Michael Spittle, Dr Paul Larkin, Mr Aden Kittel, and Dr Nathan Elsworthy.

I certify that the objectives of the study, together with any risks and safeguards associated with the procedures listed hereunder to be carried out in the research, have been fully explained to me by Mr Aden Kittel, and that I freely consent to participation involving the below mentioned procedures:

☐ Participating in video-based training (please tick).

I certify that I have had the opportunity to have any questions answered and that I understand that I can withdraw from this study at any time and that this withdrawal will not jeopardise me in any way.

I have been informed that the information I provide will be kept confidential.

Signed: _____ a

Date: _____ a

Any queries about your participation in this project may be directed to the researcher
Associate Professor Michael Spittle
Ph: 03 9919 9512

If you have any queries or complaints about the way you have been treated, you may contact the Ethics Secretary, Victoria University Human Research Ethics Committee, Office for Research, Victoria University, PO Box 14428, Melbourne, VIC, 8001, email Researchethics@vu.edu.au or phone (03) 9919 4781 or 4461.

Appendix G: Information statement (Study 4)

INFORMATION TO PARTICIPANTS INVOLVED IN RESEARCH

You are invited to participate

You are invited to participate in a research project entitled "*Decision-making training for Australian football umpires*".

This project is being conducted by a student researcher Mr Aden Kittel as part of a PhD study at Victoria University under the supervision of Assoc Prof. Michael Spittle and Dr. Paul Larkin from the College of Sport and Exercise Science Victoria University, and Dr Nathan Elsworthy from Central Queensland University.

Project explanation

Decision-making accuracy is considered the most important performance component for Australian football (AF) umpires, given the substantial impact each decision can have on the final outcome of a match. Although officiating actual match play is currently regarded as the best form of practice for umpires, there are only a finite number of games available to officiate. Therefore, it is imperative to investigate other possible training methods for umpire decision-making skill development.
To replicate game-like performance, video-based methods (i.e., watching match vision) are believed to promote the development of decision-making in officials. With the on-going development of video technology, it is also possible that 360° video virtual reality may be a more effective method to develop decision-making skills. This research will develop a video-based (i.e., match broadcast footage) and a virtual reality (i.e., 360° video virtual reality) decision-making training program. The overall aim of the project is to develop a training tool to improve match performance of Australian football umpires.

What will I be asked to do?

This study will include three decision-making tests (30 min each), and five short training sessions (10 min each). The testing sessions will require you to watch individual clips of two different video methods (i.e., 360° video virtual reality and match broadcast footage), and provide a decision as you would in a game (e.g. "play on", "holding the ball" etc.). Each test will include a 5 min familiarisation including several clips to demonstrate the type of scenarios you will be watching. Following the first test, you will be randomly assigned to one of three groups for the training program; either watching 360° video virtual reality footage, match broadcast footage, or no footage. These training sessions will be once per week (10 min each) for five consecutive weeks. One week, and one month following the training program, you will complete the decision-making test again.

The process is as follows:
- Week 0: Pre-test (360° virtual reality and match broadcast test) (30 min)
- Weeks 1-5: Training once per week (either 360° virtual reality, match broadcast, or no footage) (10 min each)
- Week 6: Post-test (360° virtual reality and match broadcast test) (30 min)
- Week 10: Retention-test (360° virtual reality and match broadcast test) (30 min)

What will I gain from participating?

It is anticipated that this project will have several benefits to you. Firstly, you will be able to experience virtual reality as a novel mode for Australian football decision-making. This research will evaluate the effectiveness of different

video-based methods in developing decision-making skill. This knowledge will assist in game decision-making of Australian football umpires to improve overall performance.

How will the information I give be used?

Results from this study will be used to enhance current Australian football umpires training practices. This knowledge will enhance match performance, training, and talent identification of Australian football umpires.

No individual data will be reported. Whilst results will be published in scientific journals, personal information will not be disclosed and all data will be coded to prevent identification of specific individuals. Specifically, only group-based averages will be used in any publications and reports emanating from this project.

What are the potential risks of participating in this project?

The associated risks of this project are minimal as the testing and training programs will be conducted at your regular training venue. Virtual reality does occasionally cause motion sickness, however, you will remain seated which should negate any symptoms. If you do not wish to participate in this study there will be no ramifications in terms of the level of service, or have any influence on current or future associations with either Victoria University or the AFL umpiring department.

How will this project be conducted?

The structure of this project is as follows:

- Week 0: Pre-test (360° virtual reality and match broadcast test) (30 min)
- Weeks 1-5: Training once per week (either 360° virtual reality, match broadcast, or no footage) (10 min each)
- Week 6: Post-test (360° virtual reality and match broadcast test) (30 min)
- Week 10: Retention-test (360° virtual reality and match broadcast test) (30 min)

This study will be administered by the student researcher (Mr Aden Kittel).

Who is conducting the study?

Assoc. Prof. Michael Spittle
Academic
Victoria University
Ph: 03 9919 9512
Email: Michael.Spittle@vu.edu.au

Mr Aden Kittel
PhD Candidate
Victoria University
Ph: 0402 636 419
Email: Aden.Kittel@live.vu.edu.au

Dr. Paul Larkin
Research Associate
Victoria University
Email. Paul.Larkin@vu.edu.au

Dr Nathan Elsworthy
Lecturer
CQUniversity
Ph: 07 4940 3382
Email: N.Elsworthy@cqu.edu.au

Any queries about your participation in this project may be directed to the Chief Investigator listed above.
If you have any queries or complaints about the way you have been treated, you may contact the Ethics Secretary, Victoria University Human Research Ethics Committee, Office for Research, Victoria University, PO Box 14428, Melbourne, VIC, 8001, email researchethics@vu.edu.au or phone (03) 9919 4781 or 4461.

Appendix H: Consent form (Study 4)

CONSENT FORM FOR PARTICIPANTS INVOLVED IN RESEARCH

INFORMATION TO PARTICIPANTS:

We would like to invite you to be a part of a study into *"Decision-making training for Australian football umpires".*

This study will involve completing decision-making testing (30 min each session) before and after a five week training program. Each video test will include two video modes (i.e. 360° virtual reality and match broadcast footage) containing 50 clips in total, where you will be asked to provide a decision on what your decision is (e.g. "play on", "holding the ball" etc.). A five week training program will immediately follow the first test, where you will be provided with decision-making scenarios once per week (10 min per session), for five weeks. This program will be followed by two more testing sessions as per the first test. The two final testing sessions will be one week, and four weeks following completion of the training program, respectively.

This study will be administered by the student researcher (Mr Aden Kittel).There are minimal risks associated with this research. While unlikely, you may feel symptoms of motion sickness while using virtual reality.

CERTIFICATION BY PARTICIPANT

I, _____ (full name)

of _____ (suburb)

CERTIFICATION BY PARENT/GUARDIAN

I, _____ (full name)

of _____ (suburb)

certify that I am giving my consent to for my child to participate in this study:
Decision-making training for Australian football umpires
being conducted at Victoria University by Associate Professor Michael Spittle, Dr Paul Larkin, Mr Aden Kittel, and Dr Nathan Elsworthy.

I certify that the objectives of the study, together with any risks and safeguards associated with the procedures listed hereunder to be carried out in the research, have been fully explained to me by Mr Aden Kittel, and that I freely consent to participation involving the below mentioned procedures:

☐ Participating in video-based training (please tick).

I certify that I have had the opportunity to have any questions answered and that I understand that I can withdraw from this study at any time and that this withdrawal will not jeopardise me in any way.

I have been informed that the information I provide will be kept confidential.

Signed: _____ a

Date: _____a

Any queries about your participation in this project may be directed to the researcher
Associate Professor Michael Spittle
Ph: 03 9919 9512

If you have any queries or complaints about the way you have been treated, you may contact the Ethics Secretary, Victoria University Human Research Ethics Committee, Office for Research, Victoria University, PO Box 14428, Melbourne, VIC, 8001, email Researchethics@vu.edu.au or phone (03) 9919 4781 or 4461.

Appendix I: Plain language information (Capture of 360° VR)

Investigating the use of 360 degree footage as an assessment and training tool for Australian football umpires

Research Team: Assc Prof Michael Spittle, Dr Nathan Elsworthy, Dr Paul Larkin, Mr Aden Kittel

Contact: Aden Kittel (0402 636 419 or aden.kittel@live.vu.edu.au)

Purpose: To identify the most appropriate viewing mode for decision-making testing and training in Australian football umpires.

Method: Scheduled training drills will be filmed from training sessions using a 360 degree camera mounted on a tripod next to the field of play, and a camera in the grandstand to recreate a broadcast viewing perspective. Drills to be filmed include small-sided games, set plays, and scratch matches. These drills will be chosen by the coaches as per normal training schedules. Following recording of this footage, it will be presented to Australian football umpires of different levels who will be asked to identify free kicks to simulate in-game decision-making. Decision-making accuracy will be assessed, followed by a training intervention to develop decision-making skill given its importance on the game of Australian football. It is anticipated approximately two months of training sessions will need to be filmed.

Results: These clips will be edited for presentation to umpires. Free kicks from the training drills will be determined by a panel of Australian football umpires coaches. The validity and reliability of this footage will be assessed.

Benefits: Recorded footage will be distributed to the participating football team being filmed. This is the first study in sports, and the first time in Australian football that 360 degree footage has been recorded with the aim of enhancing athlete performance. The footage obtained will provide a novel way to view training sessions, which is anticipated to be advantageous to team performance.

Appendix J: Plain language information (Umpire testing)

Investigating the use of 360° VR and match broadcast video as an assessment and training tool for Australian football umpires

Research Team: Mr Aden Kittel, Dr Paul Larkin, Dr Nathan Elsworthy, Assc Prof Michael Spittle

Contact: Aden Kittel (0402 636 419 or aden.kittel@live.vu.edu.au)

Purpose: To identify the suitability of using 360° video virtual reality or match broadcast footage as a decision-making tool. This tool can be used as a reliable measure for performance assessment, talent benchmarking, and identification.

Background: Decision-making assessment and training for Australian football umpires typically uses match broadcast footage to show potential free kick scenarios. With the advent of technology, virtual reality, using 360° footage, has been proposed to be an appropriate decision-making tool.

Method: Game-based training simulations of a TAC Cup team have been filmed using a 360° camera. These videos have been edited to present short clips where an infringement based decision may occur (e.g. holding the ball, high tackle, play on etc.). Similar, match broadcast footage has been edited to present clips with a potential free kick. To determine the effectiveness of each video mode, 60 clips will be presented to 10 Australian football umpires of different levels (i.e., 10 x AFL, 10 x VFL, 10 x community league/EDFL) over four occasions. Each test will take 20 min. The process is as follows:

- Session 1a: 360° video test virtual reality test 1 (20min)
- Session 1b: Match broadcast video test 1 (20min)
- Session 2a: 360° video test virtual reality test 2 (20min)
- Session 2b: Match broadcast video test 2 (20min)

It is essential each video mode is tested twice to ensure reliability (how accurately the test measures performance), and across multiple levels to ensure validity (how effectively the test differentiates across skill levels).

Benefits: This study will have the following key benefits:

- Develop a *controlled off-field* test for talent identification purposes.
- Determine the effectiveness of each video mode in *predicting on-field decision-making performance*.
- Evaluate *effectiveness of short and long-term decision-making training programs*.
- AFL umpires will be *one of the first sporting groups* to utilise this technology.

An example of the 360° video footage presented in a VR headset:

Appendix K: Plain language information (Umpire training)

Investigating the use of 360° VR and match broadcast video as an assessment and training tool for Australian football umpires

Research Team: Mr Aden Kittel, Dr Paul Larkin, Dr Nathan Elsworthy, Assc Prof Michael Spittle

Contact: Aden Kittel (0402 636 419 or aden.kittel@live.vu.edu.au)

Purpose: To identify the effectiveness of using 360° video virtual reality (VR) and match broadcast footage as a decision-making training tool.

Background: Decision-making assessment and training for Australian football umpires typically uses match broadcast footage to show potential free kick scenarios (e.g. "What's your decision?"). With the advent of technology, virtual reality, using 360° video footage has been proposed to be an appropriate decision-making tool.

Method: Game-based training simulations of a TAC Cup team have been filmed using a 360° camera. These videos have been edited to present short clips where an infringement-based decision may occur (e.g. holding the ball, high tackle, play on etc.). Similar, match broadcast footage has been edited to present clips with a potential free kick. To determine the effectiveness of each video mode, the umpires will complete a 5 week training program using one of the video methods. This training program will consist of approximately 20 individual decision-making scenarios per session, with feedback on the correct decision following each clip. The process is as follows:

Participants will be split into 3 even groups: 360° VR (Group 1), match broadcast (Group 2), normal training (Group 3).
- Week 0: Decision-making pre-test (30min).
- Weeks 1-5: Decision-making training once per week for Groups 1 & 2 (10min per session).
  Group 3 complete normal training only.
- Week 6: Decision-making post-test (30min).
- Weeks 7-9: All groups normal training only.
- Week 10: Decision-making retention-test (30min).

It is imperative that umpires in Groups 1 & 2 are at each of the training sessions in weeks 1-5.
All participants must complete the three testing sessions.

Benefits: This study will have the following key benefits:
- Determine the effectiveness of each video mode in *predicting on-field decision-making performance*.
- AFL umpires will be *one of the first sporting groups* to utilise this technology.
- Provide decision-making *training for injured athletes*.
- *Improve decision-making* of Australian football umpires.

An example of the 360° video footage presented in a VR headset:

Appendix L: Ecological validity scale (Study 2)

Make a dash on the line for how each method feels like your match decision-making process

Nothing like match decision- _____ Exactly like match decision-making

iPad

Nothing like match decision- _____ Exactly like match decision-making

Appendix M: Participant questionnaire (Study 4)

Thank you for completing the decision-making test.

Please complete the following questions which assess the suitability of VR & iPad footage.

1) Name: Date of birth:

2) How many years and games (approx) have you participated in at each level. If nil, leave blank.

	Umpired		**Played**	
Juniors	(years)	(games)	(years)	(games)
U/19	(years)	(games)	(years)	(games)
Seniors (not first grade)	(years)	(games)	(years)	(games)
First grade seniors	(years)	(games)	(years)	(games)
TAC Cup	(years)	(games)	(years)	(games)
VFL	(years)	(games)	(years)	(games)
AFL	(years)	(games)	(years)	(games)

3) Make a dash on the line for how you perceive each to be like your match decision-making process.

VR

iPad

Nothing like your match decision-making process. → Exactly like your match decision-making process.

4) Please rate VR & iPad footage on a scale of 1 (low) to 10 (high) for the following:

Relavance (how appropriate is this for umpiring AFL?)

VR: iPad:

Effort (how hard did you find the decision-making component of the task?)

VR: iPad:

Enjoyment (would you happily use this as a coaching tool?)

VR: iPad:

Concentration (how much did you need to focus?)

VR: iPad:
