



Original research

Development of a valid and reliable video-based decision-making test for Australian football umpires

Paul Larkin^{a,*}, Christopher Mesagno^a, Jason Berry^b, Michael Spittle^c^a School of Health Sciences, University of Ballarat, Australia^b Institute of Sport, Exercise and Active Living, Victoria University, Australia^c School of Exercise and Nutrition Sciences, Deakin University, Australia

ARTICLE INFO

Article history:

Received 20 February 2013

Received in revised form 29 July 2013

Accepted 2 August 2013

Available online 11 August 2013

Keywords:

Reliability

Validity

Cognition

Perception

Motor skills

Sports officials

ABSTRACT

Objectives: To develop a valid and reliable video-based decision-making test to examine and monitor the decision-making performance of Australian football umpires.

Design: Validation assessments with test re-test reliability.

Methods: A video-based decision-making test was developed from a pool of 156 video-based decision-making situations. Australian football umpires ($n = 56$) and players ($n = 45$) participated in two separate phases of analysis. In phase one, players completed a test re-test reliability assessment with a 100 video-clips. Results indicated 24 clips were a reliable measure of decision-making performance. In phase two, umpires completed a test re-test protocol with 80 clips, 24 of which were the reliable clips identified by the player cohort in phase one. These 24 clips provided a measure of construct validity. Face and content validity were assessed by skill acquisition specialists, expert umpire coaches, and umpires.

Results: From each of phase one and two of the reliability assessment, 24 clips were found to have a kappa value greater than 0.30, providing a total of 48 reliable video-clips. Construct validity was supported, with the umpire group performing significantly better than the player group on the 24 clips presented to both groups on each testing occasion. Face and content validity were also demonstrated.

Conclusions: This investigation demonstrated the ability prospectively to determine reliability and validity of the video-based decision-making test designed specifically for Australian football umpires. Establishing the validity and reliability of the video clips ensures future investigations can accurately and consistently measure the decision-making performance of Australian football umpires.

© 2013 Sports Medicine Australia. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Effective decision-making is a key component of in-game performance for both athletes¹ and umpires,^{2,3} with skilled performance characterised by consistent, efficient, correct decision-making throughout a game.¹ Numerous studies have developed methods to assess the decision-making skill of soccer^{4,5} and rugby officials,⁶ however, only one study⁷ to date has examined how to assess and/or monitor the decision-making performance of Australian football (AF) umpires.

Investigations have demonstrated the potential for sport-specific video-based decision-making tests to differentiate performance of skilled and less skilled officials.^{2,7} Video-based tests are seen as an appropriate method to assess decision-making performance because of the ability to present developing,

in-game decision-making situations, which are probably more ecologically valid than presenting static images and provide greater experimental control than in-game testing. Currently, there are no standardised decision-making assessment tools for AF umpires. Therefore, from a practical perspective, a valid and reliable video-based decision-making test may be used to identify future umpire talent, monitor decision-making development, and establish skill benchmarks for umpire selection purposes.

Despite encouraging findings for the use of video-based methods to assess the decision-making skills of sports officials,^{2,5,7} a limitation is that no umpire-based study has reported the reliability and/or validity of the video-based test. Without this assessment the test may not measure what it claims to measure, and it is not possible to determine the consistency of responses obtained.⁸ This is problematic because it is unclear whether differences in umpire decision-making performance are a result of skill differences or an unreliable test. Therefore, further empirical research is needed to determine the validity and reliability of video-based tests before the continued use of such measures.

* Corresponding author at: Faculty of Education and Social Work, University of Sydney, Sydney, New South Wales 2006, Australia.

E-mail address: paul.larkin@sydney.edu.au (P. Larkin).

New video-based decision-making tests should be based on the recognised theoretical frameworks that underpin sport-based decision-making performance. Consequently, video information must provide a decision-making situation that enables the generation of possible outcomes before selecting an appropriate decision.⁹ For a test to be specific to an umpire cohort, the video should present scenarios where participants can apply their knowledge of game rules.¹⁰ By acknowledging the theoretical underpinnings of sport-based decision-making, it is likely that the quality of decision-making tests will be improved. Therefore, the aims of the current study were to develop a video-based test to examine the decision-making performance of AF umpires and establish the validity and reliability of this test to assess decision-making performance.

2. Method

AF umpires ($n = 56$) were recruited from Division 1 competitions (i.e., local level) in metropolitan and regional Victoria (Australia). An additional group of AF players ($n = 45$) from regional Division 1 and 2 competitions in Victoria were selected to enable construct validity to be established and provide an appropriate sample size for reliability tests. All players were limited to no experience in umpiring competitive AF games. Ethics approval was granted by the University's Human Research Ethics Committee (B10-013).

To ensure face validity of the video-based test and because infringement identification was considered a key determinant of umpire performance, specific inclusion criteria were developed to enable objective selection of decision-making situations from recorded AF games. The inclusion criteria were developed following consideration of the Stages of Sport Decision-Making Model⁹ and Cornerstones Model for Refereeing Performance.¹⁰ The criteria involved two main components: (i) actual or possible rule infringement, and (ii) sufficient information to make an educated decision.

When considering face validity, the test should closely resemble the natural environment in which the skill is conducted. Therefore, for umpire decision-making, video footage should be presented from the in-game perspective of an umpire, however gaining this view is challenging due to technology and playing regulations.¹¹ As a result, video footage from a sideline or broadcast perspective is widely regarded as appropriate because it still simulates the actual playing environment, while ensuring that critical and relevant decision-making information can be identified.^{7,12,13}

Following the establishment of and consensus about the inclusion criteria by the authors, 23 games from the 2007 to 2009 Australian Football League (AFL) seasons were reviewed between two and four times, by the first author, for critical decision-making situations that met both criteria. This process identified 700 decision-making situations, which were possible infringements. These situations were edited to video-clips of 6–15 s in length using the ADOBE Premiere Pro CS4 video editing package and catalogued. This clip length was deemed sufficient to provide adequate time for the participants to orientate themselves to the context of play.^{12,14}

The first author further reviewed the 700 catalogued clips to confirm that the videos provided adequate information to enable an appropriate decision to be made. In this additional review process, the clip was only deemed appropriate if the camera angle allowed direct view of the incident. This process yielded 156 suitable test clips.

Before the commencement of each video trial, a fore-period of 2 s was provided, whereby a yellow screen with the trial number overlaid was presented. Following the fore-period, one of the 156 suitable clips was presented. Audio information was removed from the footage to ensure this did not influence participant responses. In accordance with umpire video-based research, the occlusion

method¹⁵ was used. Each video trial presented an evolving passage of game-play and occluded one second after a potential infringement. At this point, a static image of the last frame of the clip was presented for 0.5 s, followed by a blue screen with the words 'make your decision' overlaid. This indicated the commencement of an answer period of six seconds, which pilot testing indicated was sufficient time for participants to record their decision/answer. Further, as in-game decision-making occurs under pressure due to game dynamics,¹⁶ by enforcing a time pressure, the video-based test was more likely to reflect in-game decision-making. During the answer period, participants provided written responses on an answer sheet. Participants were informed prior to testing that the test aimed to replicate in-game decision-making situations, thus, participants were aware that each video-clip may, or may not, contain an infringement situation. Participants therefore, recorded whether there was an infringement or not by circling a yes/no response and then wrote their justification for the decision (i.e., a push in the back, play on no infringement), after which the justification for each decision was analysed for reliability and validity. Immediately following the answer period, a red screen appeared for two seconds to indicate the transition to the next trial. This sequence was identical for each video-clip within the test. Prior to the finalisation of the test procedure, four skill acquisition specialists reviewed and commented on the test, with all feedback addressed and appropriate modifications made to improve the test structure.

The correct outcome of each video-clip was determined by an expert panel of umpire coaches ($n = 3$, regional Victoria Division 1 competition head umpire coaches). The coaching panel had on average, 15 years' Australian football umpiring experience and 8 years of umpire coaching experience. The coaching panel was presented with the individual clips, and then independently recorded their response. All responses were collated and tallied with any discrepancies in the outcome of the clip discussed in a round table forum until consensus was reached for the clip infringement decision. For analysis purposes, the outcomes decided upon by the coaches were deemed as correct.

To ensure the video-based test adequately assessed umpire decision-making performance (content validity), an expert panel of elite/professional current umpires ($n = 7$) and umpire coaches ($n = 2$) from the AFL, attended a facilitated session to review and provide critical feedback on the test protocol. Specific feedback was noted and addressed prior to implementation to improve overall test quality. For example, four video clips within the test were replaced as the panel deemed them inappropriate due to specific AF rule changes. Furthermore, the panel stated the need for a warning slide to be presented prior to the next trial to ensure participants would be prepared for the impending trial, which was implemented in testing.

Larkin et al.⁷ found that a video-based decision-making test can differentiate skilled and less skilled Australian football umpires; however, as the current investigation developed a new test with more recent and diverse video footage, reassessment of skill level differences is required to establish the construct validity of this new test. Therefore, decision-making performance was compared between umpires (i.e., skilled) and players (i.e., players) to determine whether the video-based test could differentiate skill level differences, thus establishing construct validity.

Finally, the stability of individual responses was determined by test re-test reliability over a short time interval (14 days). This time interval adheres to Pedhazur and Pedhazur-Schmelkin's¹⁷ recommendations that the two administrations should be separated by a short interval (preferably 10–24 days) to assess random differences that characterise the test rather than changes in behaviour and ability. On both test occasions, standardised procedures and instructions were followed.

Download English Version:

<https://daneshyari.com/en/article/2704466>

Download Persian Version:

<https://daneshyari.com/article/2704466>

[Daneshyari.com](https://daneshyari.com)