Implementing the technology shift from 2D to 3D: insights and suggestions for umpire educators

Keone Kaiser, Simon Walters, Kevin Sheehy, Eoin Murray and Kirsten Spencer*

School of Sport and Recreation, Sport Performance Research Institute of NZ, Auckland University of Technology, Auckland, New Zealand

Effective methods to improve decision-making in sports officiating, particularly with the current and ongoing issues with in-person training, means new remote training methods must be developed. Traditional training of officials occurs primarily as “in-person coaching” at matches, with supporting training manuals and 2D broadcast analysis. However, live matches present difficulties in implementation, while manuals and 2D broadcast videos may not sufficiently ensure learning transfers to real situations. Due to its ease of use, first-person perspective, and ability to analyze live and post-event remotely, 360° virtual reality video technology (360° VR) offers an alternative technological solution for developing decision-making accuracy for sports officials across multiple sports. This study sought to assess the ecological validity (EV) of using 360° VR technology to enable remote teaching of decision-making. Decision-making quality and accuracy in softball umpires were similar when using 360° VR and the traditional method (2D broadcast footage), but 360° VR received significantly higher EV values than the 2D broadcast videos. In addition, interviewed participants expressed excitement about using 360° VR to augment traditional umpiring manuals and rulebooks. The results show that 360° VR has the potential to enhance or replace traditional remote learning methods for decision-making in softball umpires and allow softball learning to reach a much wider umpire audience.

KEYWORDS
softball, virtual reality, ecological validity, decision-making, declarative knowledge, procedural knowledge, umpires, remote training

Introduction

Softball is a popular bat-and-ball sport with clubs and associations around the world. Each game requires a minimum of two and preferably up to four umpires (home plate and base umpires), and with pitching speeds of over 100 km/h and fast-moving athletes, skilled umpires are at a premium, according to the World Baseball Confederation (2020b, 2020a, 2017a). Due to the speed of the game, all umpires are required to provide a “safe” or “out” decision within seconds (Szymanski and Fredrick, 2001), and with no video replays allowed to help adjudicate calls, umpires must process and execute accurate decisions quickly. However, the development of effective methods for improving decision-making in sports officiating is one of the least studied areas across all facets of sports education, and, as in most sports, the majority of decision-making training research in softball is primarily aimed at athletes and not officials (Walters et al., 2016; Kittel et al., 2019). In order to execute their
decision-making duties effectively, sports officials must learn how to align declarative knowledge (the rules of the game) with procedural knowledge (how a rule is applied to the game) in a fast-paced, dynamic environment (Plessner and MacMahon, 2013). To date, sports officials have been required and expected to master the declarative knowledge aspect of decision-making by studying manuals and passing tests or exams to demonstrate competence (Catteeuw et al., 2009; Samuel, 2017; Gulec, 2019). In the case of softball, this consists of a series of workbooks allowing umpires to progress through the qualifying levels and become qualified umpires. Procedural knowledge, however, is more difficult to teach (World Baseball, 2020b; Softball New Zealand, 2020).

Decision-making can be directly taught during games, allowing for the ideal perspective and knowledge transfer, and experience has been associated with improved ability across a range of sports (MacMahon, 2007; Catteeuw, 2009; Nazurudin, 2015). Recent research to address the limitations of traditional learning methods has focused on the use of simulated games to review plays and rules (Webb, 2014; Samuel et al., 2017; Samuel et al., 2019). While such simulations can capture the contextual and technical perspectives of actual games for the participating sports officials, the cost in manpower and time is prohibitive (Schweizer et al., 2011; Samuel, 2017; Armenteros et al., 2018). In addition, the experience of travel restrictions during the recent pandemic, which prevented in-person teaching, has resulted in a new focus on distance education and remote teaching to prevent similar issues in the future. These types of remote teaching resources will also allow for greater flexibility in time and mode of learning and reduced costs, but while they have shown similar efficacy to traditional methods, there is still much work to be done in this field (Cushion, 2018; Pérez-Camarero, 2022). The primary method for remote distance education of sports officials in recent years has been video analysis of pre-recorded or live game footage (Mascarenhas et al., 2005; Kittel et al., 2019a; Helsen et al., 2019). This allows learners to train their visual and perceptual experience of games without being present or a factor in them (Larkin et al., 2017). Remote video analysis teaching methods result in more control and consistency of training scenarios than simulation or live games, and officials can focus on key plays to improve their abilities in specific areas (Larkin et al., 2011; Larkin et al., 2015). Despite the prevalence of video analysis for sports officials decision-making training, several criticisms have accompanied its use.

While video analysis can help with routine decisions, earlier studies have found that decisions requiring dynamic, time-sensitive tasks did not seem to benefit from video analysis training methods (Edwards and Newman, 1982; Beach and Lipshitz, 1993). A more recent critique of video analysis noted that because of its 2D third-person perspective, broadcast video analysis was not representative of actual in-game decision-making demands, thereby diminishing the decision-making training potential (Araújo et al., 2006; Petit and Ripoll, 2008; Helsen et al., 2019). To date, only one study has demonstrated a transfer of sports officials decision-making ability from video match analysis to real-world performance improvement (Nazurudin et al., 2015). As such, a method of remote learning that retains the ease of use and accessibility of video learning but is more representative of actual game umpiring mechanics is essential. Virtual reality (VR) has gained traction over the past decade as a promising method of technology-enhanced learning to remotely train athletes in a range of domains (Neumann et al., 2018). In numerous sports, including baseball, VR movements were found to elicit similar responses to real-world responses (Gray, 2017; Bideau et al., 2004; Vignais et al., 2009; Kittle, 2019, 2020). VR can also manipulate variables to affect performance, both within the VR realm and the real world (Godse et al., 2019), and, importantly, VR also has advantages over traditional 2D video through its perspective presentation. Rather than seeing the game through an allocentric (third-person) view, softball umpires in VR experience the game through an egocentric (first-person) view, similar to their perspective during a game. This first-person perspective would allow an umpire to move their point of view and change their viewing angle to acquire more information as they would in a game, which may facilitate a more game-like processing of the myriad factors needed for decision-making (Craig, 2013).

The concepts of ecological validity could be applied to estimate the efficacy of a new representative learning method to improve decision-making in softball umpires. Ecological validity (EV) can be seen as the correlation/empirical relationship between a cue and its criterion and can be represented in numerical form (Brunswik, 1956). In the case of softball umpiring, an umpire may judge the accuracy of a softball call (criterion) by, for example, the speed of a runner (cue) crossing a plate. This cue would be one of many needed in order to make the call, and each would have its own ecological validity value. The strength and inter-relatedness of such cues can be estimated in order to allow an individual the chance to act on the relevant information, make a decision in a performance context, and estimate the effectiveness of an intervention (Araújo et al., 2006, 2007, 2019).

This study sought to introduce and assess the ecological validity of 360° VR technology in the remote teaching of sports official decision-making. The accuracy of decision-making and the quality of output in a cohort of softball umpires using 360° VR technology were compared to 2D broadcast footage, the current standard method of distance learning for softball umpires. The use of 360° VR as a method to differentiate the skill levels in decision-making between novice and expert umpires was also assessed (Kittel et al., 2021).

Methods

Design

A mixed-methods 2 × 2 crossover approach (Wellek and Blettner, 2012) with four progressive stages was used (Figure 1):

Stage 1 consisted of softball video clip collection in 2D broadcast and 360° video formats;
Stage 2 consisted of an expert panel formation to view and assess various video clips for study selection and inclusion;
Stage 3 required study participants to view and make a decision on softball video clips shown in 2D broadcast and 360° VR formats, and their responses were assessed according to expert advice;
Stage 4 consisted of semi-structured interviews with a selection of three umpires to obtain supplemental qualitative data and more fully understand user perspectives on the application of 360° VR as a remote training tool for softball umpires.
Practical considerations

Prior to this study, there were no existing databases of softball games filmed in 360° VR, so several games in a tournament in New Zealand were recorded in June 2020 using both a 360° VR camera (Samsung Gear 360° camera and Samsung ActionDirector 2.0) and standard 2D broadcast methods (over and above home plate), both filming second base.

The enhanced footage was obtained using a 360° VR camera with a simple mount on the umpire’s cap to reduce obtrusiveness and ensure the footage was as close as possible to the experiences of the match umpires. It was shown to participants using an Oculus Quest VR HMD unit. The broadcast footage was obtained from the broadcasters and, in all cases, was taken above and over home plate and focused on second base. The height above the ground and the quality of the footage were beyond the control of this study and, while every effort was made to ensure consistency, varied slightly from clip to clip.

The footage for each match from both the 360° VR camera and the 2D broadcast was assessed by an expert panel of umpires, and a number of plays were selected and shown in both formats to the participants according to the experimental design. Live game footage was not used in this study due to the need for the clips to be rated by the expert panel of umpires. Accuracy was assessed post-test according to the decisions of the expert panel. Live game footage was used to estimate how close their decision-making processes in each of the two video conditions felt compared to their regular in-game decision-making processes (10 being very game-like, and 0 being not at all game-like) (Catteeuw et al., 2010; Kittel et al., 2019b).

Participants

Participants were recruited in consultation with a regional softball association and were all Softball NZ-affiliated umpires with a wide range of experience and training (levels 1–7 Softball NZ Umpire training). The researcher met with these umpires in person, explained the aims of the research, and invited them to participate. The project information sheet was distributed to the umpires. The purpose of this research, risks to participants, and participant expectations were explained in detail on the information sheets.

Of the seventeen umpires who agreed to participate, two were Level 1 (rookie county level < 1-year experience), two were Level 2 (developing county level, 1–5 years experience), one was Level 3 (emerging national level, 8 years experience), one was Level 4 (badge national level, 9 years), three were Level 5 (senior badge national level, 6–11 years experience), six were Level 6 (senior NZ national level, 11–21 years experience), and two were Level 7 (certified WBSC international level, 12–37 years experience). The participants were further grouped into two broad groups—novice (Levels 1–4) and advanced (Levels 5–7) to examine whether 360° VR methods could be used to assess the differences between the decision-making skill levels of novices and experts in softball umpiring. A purposive sample of umpires was selected for further qualitative interviews, including one lower-level umpire (Level 3), one advanced-level umpire (Level 7), and one expert panel member umpire (Level 7). Online video conferencing was used with a semi-structured conversation focused on user preferences, advantages and/or disadvantages of each video condition, technological practicality, limitations, and experience level. This allowed flexibility and ease of spontaneous conversation, in addition to creating an opportunity for deeper and more individual responses from participants (MacMahon et al., 2014; Smith, 2017).

Data analysis

All quantitative data were analyzed using JASP version 0.13. Carryover effects from the treatment order (VR-2D and 2D-VR) of the 2 × 2 crossover were ruled out using a Mann–Whitney U test between the two conditions. Participant decision-making accuracy and ecological validity, as determined by an expert panel (composed of one Level 6 and two Level 7 Softball NZ umpires from across NZ), were compared between the VR and 2D broadcast conditions using a Mann–Whitney U test, and effect size was estimated using a Cohen’s test. A two-way ANOVA was used to quantify the effect of participant level and experience on accuracy. Qualitative interviews were guided by a qualitative descriptive approach (Sandelowski, 2000), and the resultant data were analyzed using an inductive thematic analysis method following (Braun and...
Clarke’s, 2006) six-step approach to further understand user perspectives on VR. An interview guide was used, with questions focused on user preferences, advantages and/or disadvantages of each video condition, technological practicality, limitations, and experience level, but open-ended questions allowed the interviewees to expand their answers as they saw fit.

Results

Quantitative analysis

A Mann–Whitney U test found no significant differences between the treatment orders (VR-2D or 2D-VR) for the number of correct calls (p = 0.523) or the ecological validity (p = 1.000), indicating no carryover effects of the 2 × 2 crossover design that would influence the outcomes of the trial.

The decision-making of the umpires was assessed by an expert panel across 10 decisions. The panel examined all of the footage and came to a consensus about the correct decision in each case, including the reasoning behind each decision. The decisions made by the umpires in the trial were assessed against this consensus. The mean ± SD for the number of correct calls for 360° VR (out of a maximum of 10) was 8.0 ± 0.9, while for 2D broadcast, they were 8.5 ± 1.5 across all umpires. The data for each treatment group, along with the distribution and median, are shown in Figure 2.

A Mann–Whitney U test indicated no significant difference in decision-making accuracy between 360° VR and 2D broadcast video users (W = 178.0, p = 0.242, r = 0.204). Furthermore, there was no significant difference between advanced and novice umpires. The mean ± SD for the number of correct calls for novice umpires using 2D broadcast videos was 8.5 ± 1.4, and 8.0 ± 1.0 for the 360° VR condition. For advanced umpires, the means for 2D broadcast and 360° VR were 8.5 ± 1.7 and 8.0 ± 0.8, respectively. The results of the two-way ANOVA test comparing the two treatment groups and the interaction between the treatment group and experience level indicate no statistically significant difference in decision-making accuracy between advanced and novice umpires (F = 0.002, p = 0.961, ηp² < 0.001), and this did not vary by video condition (p = 0.961).

Ecological validity, on the other hand, was found to be significantly different between 2D broadcast and 360° VR. The mean ± SD ecological validity score (1 low–10 high) was 7.1 ± 2.0 for 360° VR, while the mean for 2D broadcast was 4.4 ± 1.9. A Mann–Whitney U test indicated that there was a significant difference in ecological validity score between 360° VR and 2D broadcast video users (W = 32.5, p < 0.001, r = −0.626), with a moderate Cohen’s effect. The data for each treatment group, along with the distributions and medians, are shown in Figure 3.

Qualitative analysis

A purposive sample of three participants was selected for further qualitative interviews: one lower-level umpire (Level 3), one advanced-level umpire (Level 7), and one expert panel member umpire (Level 7). These discussions suggested five areas of subjective advantage for VR. These were perspective, engagement, information, enjoyment, and technological suitability and impact.

A common theme among all three interviewees was the importance of perspective in softball umpire decision-making. The participants felt a disconnect when using broadcast video, as the footage above and behind first base was not from an in-game, umpire-relevant perspective.

“‘You’re in a different angle, so you’re not utilizing techniques that you would have to from the position of the umpire.”
However, they felt more connected and confident in their decisions when using 360° VR techniques, as the perspective was from a more familiar position. The power to interact with the 360° VR footage, including the ability to change view and focus on pertinent angles, increased participant engagement and allowed increased information to be gleaned from the footage, with the most experienced umpire stating that being able to see a play through the “eyes of an umpire” helped them figure out what an umpire was thinking and if they were in the best place to make a call. Interviewees also remarked upon 360° VR’s technological suitability, particularly for younger umpires, and the draw of using a novel and innovative technique that made using it more interesting and engaging than traditional broadcast footage learning methods:

“…you can act out what you should be doing and what you would be doing at the same time.” “You look at our younger umpires … they’re technology savvy. So VR to them would be like playing a game.”

They also noted that it would be suitable for multiple learners under direct tuition, as immediate, actionable Instructions could be given to multiple people simultaneously to observe specific points and foci, including the ability to track a ball as it moved across the field from its point of origin. However, there were some subjective drawbacks regarding footage and technological suitability. One was an issue with the lack of 360° footage and that collecting it when attached to an umpire could expose the umpire to criticism for a perceived mistake or that the umpire recording may not be focused on the correct thing during the required key points of the match. There could also be a minor discomfort caused by the motion of the VR experience, leading to “VR sickness” (MacQuarrie and Steed, 2017). However, this was not cited as a major impediment to decision-making in this study.

Discussion

Previous research in other sports has highlighted the ability of both 2D broadcast and 360° VR video to improve sports official decision-making accuracy (e.g., Catteuw et al., 2010; Schweizer et al., 2011; Helsen et al., 2019; Kittle et al., 2019, 2020), and this study directly compares the two techniques of technology-enhanced learning (TEL) in softball umpiring. In this study, no significant difference was observed between the accuracy of calls made while using the 2D broadcast and 360° VR footage. With 2D broadcast footage methods historically being the primary method of remote learning and being familiar to all baseball umpires, it is very promising that umpires were able to achieve similar accuracy in calls with minimal experience of using 360° VR.

Both methods also provided no significant difference in decision accuracy across experience levels, despite interviewees generally being of the opinion that 360° VR was more applicable and usable for younger people, indicating that both forms of TEL could be applicable across the population spectrum.

Participants noted different pros and cons with both forms, with the 2D broadcast providing a better overall picture of the entire softball diamond and game contextual information while sometimes sacrificing a closer and clearer view of the desired location (second base). 360° VR, on the other hand, afforded a familiar, applicable viewing angle and perspective for umpires, but with issues around reduced resolution and extra non-stabilized motion that may be distracting when making close decisions.

Ecological validity correlates the test performance with behaviors and experiences in real-world settings (Araújo, 2006, 2007; Kittel, 2019, 2020) and was significantly higher for 360° VR than 2D broadcast. Interviewees also observed that the experience of using 360° VR allowed the umpires to view each play from a viewpoint almost identical to the real game, therefore enabling them to more easily obtain the four elements (leading edge of base/plate, ball position, fielder location, and runner location) required to make an accurate call. This increased the decision-making link between cue and criterion (Brunswik, 1956). They also highlighted the ability to “identify key triggers,” describing the experience as “immersive” and having a heightened “sense of presence,” and the resulting heightened embodiment may lead to an increasing probability of long-term behavioral changes (Bohll et al., 2011). These are all ecological validity improvements expected from the shift from the third-person of 2D broadcast footage to a more applicable first-person perspective (Croft and Riberiro, 2013). 360° VR has additional advantages of an increased field of view and interactivity, such as the ability to change the view, focus on pertinent angles, and track contextual cues that may not be possible with other first-person perspective methods.

Limitations

While prices have decreased and availability has increased, 360° VR remains expensive and a niche technology. In addition, there is a lack of suitable footage, and with the possible stigma of individual attention on calls and the techniques that recording such footage brings, future footage may be difficult to obtain. The limitations of this particular study are the relatively limited population used for qualitative research; increasing the number of participants and geographic spread would improve the accuracy and depth of the resulting data.

Conclusion

Due to its ease of use, first-person perspective, and ability to analyze live and post-event remotely, 360° VR may offer an ideal technological solution for learning decision-making accuracy for sports officials across multiple sports. The accuracy of decisions made in the traditionally successful 2D broadcast video learning methods and the 360° VR videos in softball umpiring were very similar, and 360° VR videos received significantly higher EV values than the more traditional video learning technique. Interview participants also expressed excitement about the prospect of using novel 360° VR technology to augment or replace traditional remote methods for teaching decision-making to softball umpires and allow softball learning to reach a much wider audience of umpires or even other sports.
Data availability statement

The datasets presented in this article are not readily available. Requests to access the datasets should be directed to kspencer@aut.ac.nz.

Ethics statement

This study involving humans was approved by the Auckland University of Technology Ethics Committee. The studies were conducted in accordance with local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

KK: writing–original draft, writing–review and editing, conceptualization, data curation, formal analysis, investigation, and methodology. SW: writing, data curation, supervision, writing–review and editing. EM: writing–review and editing. KiS: conceptualization, supervision, writing–original draft, and writing–review and editing.

References


Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declared that they were editorial board members of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, or the reviewers. Any product that may be evaluated in this article, or any claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.


