INTERNATIONAL JOURNAL OF SPORT AND EXERCISE PSYCHOLOGY
2024, VOL. 22, NO. 2, 382–406
https://doi.org/10.1080/1612197X.2023.2262483

Assessing the validity and reliability of a baseball pitch discrimination online task

Georgia Grieve[a], Zachary Besler[a], Sean Müller[b], Miriam Spering[c,d] and Nicola J. Hodges[a]

[a]School of Kinesiology, University of British Columbia, Vancouver, Canada; [b]Centre for Smart Analytics, Federation University, Ballarat, Australia; [c]Department of Ophthalmology & Visual Sciences, Faculty of Medicine, University of British Columbia, Vancouver, Canada; [d]Djavad Mowafaghain Centre for Brain Health, University of British Columbia, Vancouver, Canada

ABSTRACT

There has been an increasing interest in training perceptual skills in sports through online video-based methods, particularly in baseball. However, there is little empirical evidence related to the reliability and validity of such online methods for the assessment of these skill. Here we developed an online task to assess pitch discrimination and evaluated (a) inter-item reliability, (b) reliability in assessment compared to an in-person task, also tapping into external validity and (c) discriminability across different skill groups. We also compared performance on a non-sport specific Dynamic Visual Acuity task (DVA), thought to tap into underlying visual skills comprising pitch discrimination. Skilled, Varsity-level baseball players ($n = 17$) were compared to novices ($n = 14$) when discriminating pitches thrown by two different pitchers, across three pitch types, edited to progressively remove sections of ball flight (3 time points). The online task discriminated across skill groups, showed good reliability across repeated viewings and from the online task to an in-person assessment of skilled athletes ($n = 8$). There were, however, differences in reliability and discriminant validity based on the type of pitcher, with one pitcher being responded to more accurately and reliably. Skilled participants showed good discriminability between fastballs and change-ups. There were no group differences for DVA, nor did it correlate with pitch discrimination for the skilled group. These data illustrate the reliability of online video assessments, but raise issues concerning discriminability across different pitchers and when standing ready to swing. Greater sensitivity testing of such assessments is still needed, within and across skill groups.

ARTICLE HISTORY

Received 17 February 2023
Accepted 20 September 2023

KEYWORDS

Anticipation; decision making, test validity; perceptual-cognitive skills

Perceptual-cognitive skills have a long history of study in sports and are defined as skills that guide the use of environmental information, underpinning action anticipation, selection and execution (see Hodges et al., 2021; Williams et al., 2011, for reviews). They have sometimes been defined as abilities, rather than skills, as there is a continuing debate as to how much these skills are dependent on experience versus more innate ability. This is particularly true for visual "skills", such as static and dynamic visual acuity (Hodges et al.,

CONTACT Nicola J. Hodges nicola.hodges@ubc.ca

2021). In baseball, perceptual-cognitive skills have primarily been studied with respect to three types of information use: situational or contextual cues (e.g., pitch count; Gray & Canal-Bruland, 2018), body-related kinematic cues (e.g., wrist position indicating pitch type, Müller et al., 2017), and ball-flight related cues (e.g., visual gaze of ball flight; Fooken & Spering, 2020). The ability to use these sources of information has an impact on the success of a player at-bat (e.g., Müller & Fadde, 2016).

Research in baseball has shown that experts and near experts are able to make an accurate prediction about pitch type before ball release or very early in ball flight and can then decide about the location of this pitch shortly after ball release (e.g., Bahill & Karnavas, 1993; Fadde, 2006; Gray, 2010; Morris-Binelli et al., 2018). A hitter who relies just on ball flight to respond, will have limited time available to execute their response to a pitch. Thus, the earlier a batter can identify the pitch type and direction, the sooner they can initiate their swing, or decide not to swing. Research in various interception sports has shown that experts are able to use advance cues from the body before ball release and in the early stages of ball flight to guide their response (e.g., Abernethy, 1996; Fadde, 2006; Gray, 2010).

There has been a relatively recent interest in using virtual reality, virtual environments, and video game-type tools to train perceptual skills in athletes (e.g., Drew, 2021; Gray, 2017; see Müller et al., 2023 for further discussion). The COVID-19 pandemic spurred the development of online environments for physical and motor skills assessment in public health and psychological research settings (e.g., Gustafsson et al., 2021; Scott-Andrews et al., 2022) and motivated our own research of such online assessments in sports. Indeed, the research has been lagging behind the practice, given the commercialisation of online video assessment and training platforms such as gamesense (https://gamesensesports.com/baseball/), which have already infiltrated professional markets (https://www.sportsbusinessjournal.com/Daily/Issues/2022/12/09/Technology/mlb-winter-meetings-baseball-technology-trajekt-gamesense-winreality). More broadly, there are no evidence-based guidelines regarding when and how to best introduce online video-based assessments and training and limited evidence to show transfer from playing video-type games online, to performing in competitive on-field situations (Fadde, 2006; Gray, 2017; Morris-Binelli & Müller, 2017; for a review, see Gray, 2019). In this research study, our aims were to design and assess the reliability and validity of an online task for the assessment of perceptual cognitive skills in baseball players. Our focus was on pitch discrimination as one aspect of a perceptual-cognitive skill needed in baseball. Pitch discrimination refers to a batter's capability to determine, for example, the type or location of a pitch, such as whether a pitch is a fastball or curveball and whether it will be thrown inside or outside the strike zone (Müller et al., 2017). We also tested athletes on a non-sport specific test of Dynamic Visual Acuity (DVA), as a potential component skill suggested to underlie pitch discrimination (e.g., Laby et al., 2019; Uchida et al., 2013).

To assess perceptual-cognitive skills, particularly anticipatory skills related to outcome discrimination, temporal occlusion methods have been employed (for reviews, see Abernethy et al., 1994; Müller & Abernethy, 2012; Smeeton et al., 2019). An action is filmed and edited at particular points to withhold or provide information deemed important to anticipation. Elevated prediction at an occlusion point, or an improvement in prediction accuracy between occlusion points, is considered evidence of information becoming available for use (e.g., Farrow et al., 2005). In studies using temporal occlusion for assessment, expert athletes have been shown to be consistently better able to anticipate outcomes than novices and make use of advanced kinematic information from the body of an opponent, which goes undetected by novices (e.g., Abernethy & Russell, 1987a, 1987b; Farrow et al., 2005; Aglioti et al., 2008; Farrow & Reid, 2012; Müller et al., 2017). Differences in prediction accuracy between skilled and less skilled athletes have been replicated in many sports and athlete groups, particularly in sports requiring interceptive actions (see Hodges et al., 2021).

In striking sports, highly skilled, near to expert players, are significantly better at predicting the outcome of an opponent's action before ball release in bowling, serving, and pitching actions (e.g., Brenton et al., 2016; Chen et al., 2017; Farrow & Reid, 2012; Müller et al., 2006, 2017; Weissensteiner et al., 2008). Highly skilled players were able to predict outcomes above chance at the point of ball release, whereas novices were only able to predict above chance when ball flight was visible (Brenton et al., 2016; Müller et al., 2006). These results indicate that lesser-skilled players and novices rely on ball flight to make predictions, rather than body-related kinematic cues from the pitcher/thrower.

When assessing visual anticipation in professional baseball batters, Müller and colleagues (2017) showed that whereas accurate predictions could be made about pitch type early in the action, more time and ball flight were required to accurately predict the final location of the ball (see also Morris-Binelli et al., 2018). Batters completed a task requiring them to discriminate pitch type and estimate ball location based on occluded in-game footage of professional pitchers. At the first three occlusion points (ball release until 200 ms after), pitch type predictions were more accurate than pitch location predictions. This pitch-type advantage did not change until all ball flight information became available to the batter (no occlusion condition). Further, predictions were less accurate at all occlusion points when both type and location predictions were made, rather than either location or type individually. Asking batters to decide location before they have determined pitch type might interfere with their ability to accurately predict the ball's flight. Because pitch location accuracy was not above guessing level until 200 ms after ball release, it appears likely that batters are deciding about pitch type first, with professional batters able to accurately predict pitch type at ball release.

In the literature discussed so far, we have focused upon discrimination of visual anticipation using temporal occlusion perceptual tasks, but it is important to consider how these findings relate to action. Vision occlusion glasses that can create temporal occlusion in field-settings (Milgram, 1987) and chronometric analysis that can map visual-perceptual information to action responses (Abernethy, 1984), have been used to understand visual anticipation in perception-action coupled tasks. In studies using occlusion glasses to temporally occlude a performer's vision of an opponent's action before ball flight, more expert athletes are superior to lesser-skilled players at using advance visual cues for positioning the body for object interception (e.g., in tennis return of serve, Farrow & Abernethy, 2003 and cricket batting, Müller et al., 2009). Temporal occlusion manipulations during ball flight have also shown that this information is used to guide interception (e.g., in cricket batting, Müller et al., 2009). Through chronometric analysis, major league and national youth baseball batters have been shown to initiate their lead foot stride relative to the point of ball release by the pitcher, indicating use of advance information to guide body positioning for bat-ball interception (Hubbard & Seng, 1954; Canãl-Bruland et al., 2015). Bat or racket initiation occurs during ball flight (Hubbard & Seng, 1954; Canãl-Bruland et al., 2015). Accordingly, perception-action coupled field studies underscore the validity of determining advance cue usage and object flight information for visual anticipation reported in video-based tasks. Further support for this conclusion is provided by a recent meta-analysis that reported visual-perceptual information to be the main predictor of expertise discrimination in sport (Kalén et al., 2021). Therefore, understanding the validity and reliability of perceptual-cognitive skill in video-based online tasks is highly relevant to action in sport.

In baseball, hitters must respond to an object which is both low contrast and thrown from far away, while under significant time constraints (Laby et al., 2019). These visual demands have led researchers to study general visual skills, which are not specific to the sport (see Hodges et al., 2021 for review). Laby and colleagues (2019) used a Landolt C task to assess DVA in baseball players, in which the difficulty of the task was manipulated by changing the size (i.e., the letter C had different sized openings), contrast, and viewing time of the target. DVA was positively correlated with the baseball statistic of walk rate (and negatively correlated with swing rate), indicating greater selectiveness or discrimination at the plate when deciding whether to swing (Laby et al., 2019; see also Burris et al., 2018). The DVA task was also shown to discriminate between baseball players and non-players (Uchida et al., 2013). Differences in tracking ability between groups is thought to be due to differences in exposure to a specific stimulus (i.e., tracking a baseball), rather than an innate ability to move their eyes well (Palidis et al., 2017). As such, if such a general tracking skill is important to pitch discrimination, we would expect a correlation between a measure of DVA and pitch discrimination, potentially attesting to construct validity.

In this research, we developed an *online* experimental task to test pitch discrimination in a group of skilled baseball athletes. Our aims were to determine; (a) the internal and external validity of this online task by comparing across different skill groups and comparing performance to an in-person task and a non-sport specific test of DVA, as well as to assess (b) reliability; by comparing responses to repeat viewings of the same stimuli and of the same pitch delivered by different pitchers. We expected that the more skilled athlete groups would be better able to discriminate pitch type than lesser skilled athletes and that they would also show better DVA on a non-sport task, if a general ability to track fast moving objects is related to skill in baseball. A correlation between these two measures would be expected. Evidence for both reliability and external validity, as evidenced in the latter case through comparisons of the online test of pitch discrimination to an in-person assessment, were also predicted. As to the scope of this work, we developed this online test with the long-term goal of being able to collect longitudinal and cross-sectional data on perceptual-cognitive skill development of athletes in baseball, with the aim of determining sensitivity of these measures to practice experiences across development.

Methods

Participants

Male skilled baseball players ($n = 17$) were recruited via the University of British Columbia Baseball Centre and their affiliate organisations. Skilled baseball players (M age = 19.7, SD

= 1.4 yr) were defined as having more than three seasons of playing experience with either a national ($n = 4$), provincial ($n = 10$) or regional representative team ($n = 3$) in the Premier Baseball League (PBL) or in the BC Minor Baseball League. All were part of a team competing in the NAIA (National Ass. of Intercollegiate Athletics) Cascade Conference (Northwestern USA College League), thus not only of a high level of skill, but also representative of college-level athletes who would be using similar types of online software for skill development. Baseball experience information was confirmed via a customised Qualtrics' survey designed to assess an athlete's experience in terms of number of years played ($M = 14.5$, $SD = 2.04$ yr) and level of competition. The skilled baseball athletes reported participating in informal baseball practice since a mean age of 4.7 yr ($SD = 1.9$) and in coach-led organised practice since a mean age of 7 yr ($SD = 2.4$). They reported practicing a minimum of 6 hr/wk in the 2019 season (the last full season before COVID-19 related restrictions and participation in our study), but more than half of the skilled sample reported practicing more than 12 hr/wk during this time. We also recruited a group of male novice participants ($n = 14$; M age = 20.6, $SD = 2.2$ yr), who had little to no baseball experience (i.e., novice group) and hence did not fill in the baseball experience survey. Novices were recruited via the UBC Psychology Paid Participants' Studies list and classroom advertisements. We originally had collected data from n = 18 novices to match the skilled group but on later inspection of the data, four individuals were excluded because of validity issues with the responses (i.e., a lack of variation in responses to the pitch discrimination task suggesting the participants had failed to adhere to the task instructions). We conducted post-hoc power analyses in G*Power (Faul et al., 2009; Version 3.1.9.7) for both 3- and 4-way mixed-design ANOVAs. Effect sizes were based on conservative estimates from the baseball study of Chen and colleagues (2017; $n = 15$ per skill group), which yielded, at the high end, a total sample size of $N = 16$ for our study ($\alpha = .05$, 80% power, $f = .25$).

All potential participants were required to confirm their age and experience with baseball via email before they were sent the link to the experience questionnaire. Information provided in the Qualtrics' questionnaire was checked before any further invite to continue with the online experiment. All study procedures were conducted according to the guidelines of the University of British Columbia Behavioural Research Ethics' Board [H19-02705] and participants provided informed consent before participating. Participants received $15 Amazon gift card as remuneration on completion of all components of testing. A sub-sample of the Skilled athletes (n = 8) also completed the pitch discrimination task a second time in-person at the UBC baseball centre as part of pre-season baseline assessment. Informed consent obtained from a third-party researcher not involved in the study was again obtained from these athletes for secondary use of their data from the in-person assessments and to cross-validate these with the online assessments. No remuneration was provided for the in-person testing.

Stimuli

Three college-level baseball pitchers were filmed with consent to create the visual stimuli for the experiment. One pitcher's video was used for familiarisation only (he was familiar to other athletes) and the other two were chosen for the experiment as they were not familiar to the participants and had both been previous University or College pitchers.

One pitcher was recruited by an NCAA Division 1 University within two seasons of filming. The pitching speeds of the stimuli used in the experiment are given in Table 1, where it can be seen that the fastball pitches from both pitcher A (~89mph) and B (~85mph) were both fast and comparable to average speeds of players from the top tier NCAA Division I (~85mph) and Division II pitchers (80mph) (https://www.baseballmode.com/average-pitch-speed-by-age/; retrieved July 2023). Two pitchers were used in the experiment because batters commonly face more than one pitcher whilst also replicating methods used in related research (Chen et al., 2017).

A portable pitching monitor (Rapsodo PITCHING 2.0), a unit which provides information about ball release and flight during pitches, was used to confirm pitch type, final location of the pitch relative to the strike zone, and ball velocity (Aucoin, 2019; Rapsodo, 2019). Set-up of the Rapsodo unit relative to the camera and pitcher is shown in Figure 1. Pitchers were asked to throw a series of pitches towards the strike zone of a hitter who would be approximately 175–180 cm tall. The supervising coach of the pitcher, who offered to help with the filming and ensure that the pitcher did not throw more than 24 pitches because of concerns about injury/load stress, instructed the pitcher as to what pitch to throw. He also confirmed, in conjunction with the catcher and the Rapsodo output, the type of pitch and final location relative to the strike zone (i.e., ball or strike). A Sony digital video camera (frame rate 33 Hz) was set up 354 cm back from the centre of the home plate, and 64 cm to the left, with the lens at a height of 155 cm (see Figure 1). Camera and pitcher set up were based on the methods of Chen and colleagues (2017), as this simulates the perspective of an average height adult batter and ensures depth-cues are not lost by footage being taken from immediately in front of the pitcher.

Six videos were used from each of the two pitchers for the experimental stimuli, representing three different pitch types (fastball, curve ball and change-up), such that there were two videos per pitcher for each pitch type (see Table 1 for speeds). These videos were selected from amongst the 24 pitches, based on pitches that were expected to hit or be close to the strike zone. Using Windows' Video Editor, individual video clips were edited such that the outcome of the throw was occluded, with edits at various points in the unfolding action. The first occlusion point (OP1) was 133 ms after the ball left the pitcher's hand. As previous research has reported that a minimum latency for adjusting ongoing complex striking movements is in the order of 180 ms (McLeod, 1987), we expected batters to still be processing ball release information within this temporal occlusion condition. A duration of an additional 133 ms was applied to the edited video to create the next occlusion point (OP2) at 266 ms after ball release, and another 100 ms later at 366 ms for the final occlusion point (OP3). These time periods were based on the work of Chen and colleagues (2017) and corresponded to early, mid, and late ball flight conditions. The order of trials here was arranged specifically to make

Table 1. Ball speeds for pitcher A and B, for the two pitches thrown for each pitch type (fastball, change-up or curveball).

	Fastball		**Change-up**		**Curveball**	
	1	2	1	2	1	2
Pitcher A	89 mph	89 mph	75 mph	78 mph	74 mph	77 mph
Pitcher B	84 mph	86 mph	70 mph	72 mph	73 mph	73 mph

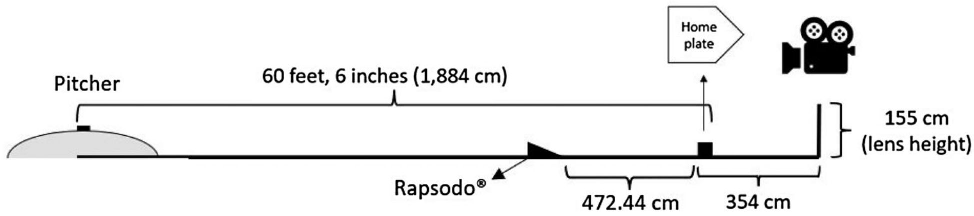

Figure 1. Set up for stimuli filming (from side-view). The gray bump on the left represents the pitcher's mound, where the pitcher pitches the ball. The small black box represents the central location of home plate for which the distances are shown relative to the pitcher's mound and to positioning of the Rapsodo unit (to capture pitch-type statistics) and to the video camera for filming of experimental stimuli (i.e., pitching and ball flight).

sure that the 366 ms clip of a given pitch was not shown before a 133 ms clip of the same pitch (potentially providing feedback on ball flight and pitch-type).

Procedure

Online test

Pitch Discrimination: After completing the Qualtrics' survey to help delineate across skill and experience and determine an individual's validity to continue, participants were sent an individual code and link to the online task (to be completed on a laptop or desktop computer, not a phone). The online experiment was conducted using the Gorilla software platform for both design and administration (https://gorilla.sc/support). Participants were asked to conduct the experiment while in a seated position as they made decisions in response to the videos and to place the centre of the screen at eye level. To facilitate this procedure and alignment, we used a feature in Gorilla where participants were asked to put a standard size credit card on the screen and to drag a slider such that the image on the screen matched the size of the card. As such, when participants were seated 50 cm from the screen, the stimuli would subtend the same visual angle, irrespective of screen size.

Participants were asked to view a series of occluded clips of either baseball pitcher and to discriminate pitch type by pressing an assigned key on the keyboard. They were encouraged to make fast responses whilst not sacrificing accuracy. Each participant saw videos of the two pitchers and three pitch types, with two different pitches for each pitch type. There were also three occlusion times and each video was repeated twice so we could measure reliability, resulting in a total of 72 pitches. Note that participants were asked to make decisions about pitch type and not location (*cf* Chen et al., 2017). Because of variations associated with screen size and projection angle as a result of online testing (and hence determination of the strike zone), we did not take measures of anticipated ball location (i.e., "in" or "out" of the strike zone). Participants were given 3 s to respond to a given clip before the response screen timed out. This time-out happened 11 times, a maximum of once/person (2 skilled, 9 novice). These trials were subsequently excluded and not repeated.

Familiarisation practice trials (t = 9), representing a selection of early, mid, and late OPs for the three different pitch types were included to help participants understand the task requirements. During the familiarisation phrase, feedback was provided as to the various types of pitches. Participants also viewed instructions to explain the differences between the three pitch types. We further included motivational feedback to encourage participants during the familiarisation and experimental trials; including "good, keep going", "well done, you're doing great" etc., that was independent of accuracy.

During the experimental phase, participants responded to the video clips in the same manner as during the familiarisation trials, inputting their predictions with a key-press corresponding to the pitch type (i.e., fastball, curveball, change-up). Each participant saw every clip (e.g., Pitcher A, fastball1, 266 ms) twice throughout the online task (72 trials total). Trials were grouped in short blocks of six trials each. Breaks were scheduled into the programme between blocks and the participant was able to control the break time. We showed clips of the same pitcher in blocks but the order of clips and pitchers was randomised across participants. The total time for the pitch discrimination task to be completed was ~15 min, depending on the amount of rest between blocks. The chronology of events in the experimental task is illustrated in Figure 2.

Dynamic visual acuity (DVA): after completing the pitch discrimination task, participants completed a short (~10 min) DVA task using the same Gorilla software platform. We used a custom-programmed moving Landolt C task, which required participants to determine the location of the opening of a rotated letter "C" which moved across the screen at high speed (approximating baseball speed), with gap sizes ranging from one pixel (hardest) to four pixels (easiest: Palidis et al., 2017; see also Laby et al., 2019 & Uchida et al., 2013). Participants responded by pressing a key signalling the location of the opening; either left-up, left-down, right-up or right-down. The same time-out period (3000 ms) as in the pitch discrimination task was used for this task, but no trials had to be excluded for time-outs.

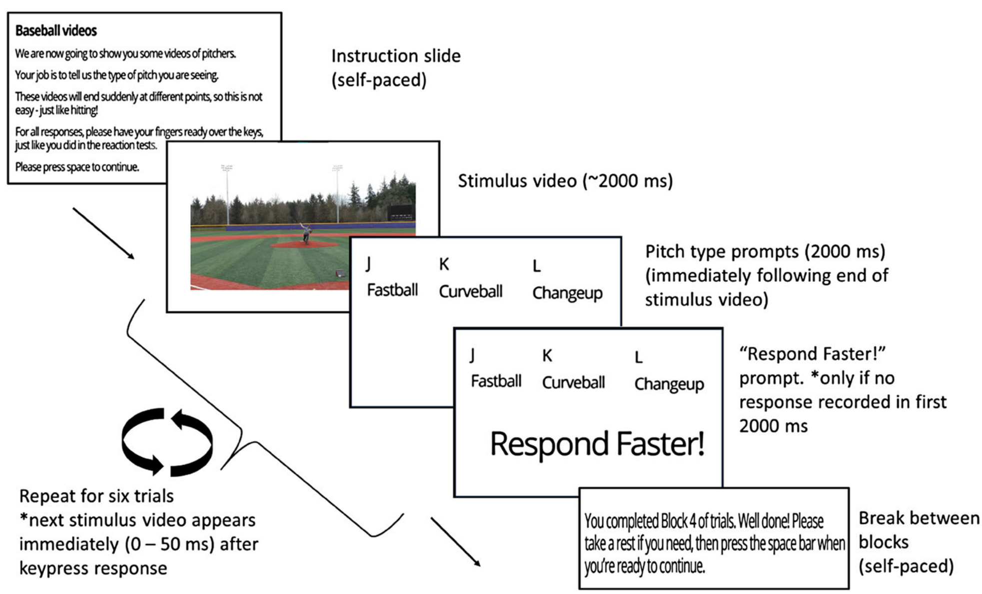

Figure 2. Chronological sequence of events for the online experimental task.

In-person test

A subset of skilled participants completed an in-person version of the task after the online task. The testing was completed at the university in-door batting cages. Participants stood in front of a flat screen monitor at a distance of ∼180 cm/6 ft (Dell P2319H, 22.5″ screen size, 1920 × 1080 resolution) with a bat, as though they were going to hit the incoming pitch (i.e., ready to swing stance). This distance was calibrated to ensure the visual angle subtended by the pitcher on the screen (∼5°) matched that experienced in typical game play with the pitcher's shoulder (i.e., screen centre) lining up with eye height (replicating previous methods, e.g., Brenton et al., 2016; Chen et al., 2017). Athletes were not, however, required to swing, but instead were asked to respond verbally, as quickly as possible, with their pitch-type response. Only discrimination accuracy and not speed was recorded. All other procedures were matched to the online task.

Data analysis

We collected data on response time and accuracy (proportion correct responses) for the online pitch discrimination task and for accuracy only for the DVA task and in-person pitch discrimination task. We compared accuracy and response time across groups and conditions and calculated perceptual sensitivity using d prime for pitch discrimination for the online task. The measure of d prime was based on comparisons of fastballs and change-ups only, where hits (correct-detections) and false-alarms (incorrect detections) were calculated for each individual. Because our task allowed for extreme hit and false alarm results (0 and 1), we used a log linear normalisation procedure (Hautus, 1995) to adjust for biases. The loglinear procedure requires that all values be corrected and so we added 0.5 to every hit and false alarm score, and added 1 to the total number of trials (Hautus, 1995). These corrected values were then entered into the standard d Prime calculation, $d' = Z(H) - Z(F)$. We chose to compare only fastballs and change-ups, excluding curveballs, as these two pitches are the most similar kinematically until late in ball flight and most difficult to differentiate.

Statistical analyses were conducted using JASP (version 0.12.2, JASP team 2020) and involved comparisons across the two skill groups when possible using mixed-design ANOVAs. For pitch discrimination, the repeated measures' (RM) variables were pitcher (A or B), pitch-type (fastball, change-up or curveball) and occlusion point (OP1-3, corresponding to release point +133, + 266 or +366 ms). We also ran a fully RM-ANOVA comparing across the online and in-person task for the subset of skilled participants who completed both tasks, with task as an additional factor. For the DVA task, gap size of the Landolt C (1-4 pixels) was the only RM factor. When comparisons involved more than two means, Tukey HSD post-hoc comparisons were applied. Cohen's *d* is presented as a measure of effect size when comparing two means and partial eta squared is reported for ANOVAs.

Correlational analyses (Pearson r) were used to relate measures of prediction accuracy to measures of DVA in order to determine whether performance in one was related to performance in another, speaking perhaps to similar underlying skills for both. We also correlated response times and accuracy for the online prediction task to evaluate any potential speed-for-accuracy trade-offs.

Test-retest reliability analysis (i.e., within-task) was conducted separately for the online task and then for the subset of athletes who completed both the in-person and online versions of the task. Because of the categorical nature of individual responses, these analyses were based on calculation of percent agreement for the same clip shown twice within a task. The percent agreement of responses to both exposures of each video clip was calculated. If participants gave the same response to both exposures, agreement was 100%, if different responses were given, agreement was 0%. Agreement was calculated at each occlusion point and then averaged across all occlusion points for each participant for statistical analyses.

To give a measure of between-task reliability, we also compared response data in the online task to that of the in-person task for the same clip. If a person had 100% agreement for the same clip on the online task, their responses to the in-person version would lead to a percent agreement of 0%, if no response matched, 50%, if one response matched, or 100% if both responses matched. If the responses to the same clip in the online task were different (i.e., 0% within-task agreement), but the individual gave these same responses again in-person, they got a score of 100% between-task agreement. This reduced to 50% if both responses were the same for the in-person version (i.e., one different from the online) or 0% if neither response matched. Agreement was calculated at each occlusion point and then averaged across all occlusion points for each participant for statistical analyses.

Results

Pitch discrimination: online task

Proportional accuracy

Skilled athletes scored more accurately ($M = .58$, $SD = .34$) than the novice participants ($M = .33$, $SD = .30$) on the pitch discrimination task, as confirmed by a main effect of skill group, $F(1,29) = 29.04$, $p < .001$, $\eta^2_p = .50$. A single sample t-test showed that only the skilled group performed statistically above chance (>.33; $p < .001$). Participants responded differently to the different pitchers as illustrated in Figure 3. The proportion of correct responses was higher for pitcher B ($M = .54$, $SD = .35$) than for pitcher A ($M = .39$, $SD = .32$), $F(1,29) = 39.65$, $p < .001$, $\eta^2_p = .58$. We also observed main effects for pitch-type, $F(2,58) = 35.88$, $p < .001$, $\eta^2_p = .55$ and occlusion point (OP), $F(2,58) = 3.79$, $p < .05$, $\eta^2_p = .12$. The fastball and curveball were responded to more accurately than the change-up (both $ps < .01$) and OP3 (+366 ms) was responded to more accurately than OP1 (+133 ms), $p < .05$. Skill group did not interact with Pitcher, $F(1,29) = 2.32$, $p = .14$, $\eta^2_p = .07$, nor OP, $F(2,58) = 1.91$, $p = .16$, $\eta^2_p = .06$, but it did with Pitch-type, $F(2,58) = 5.16$, $p < .01$, $\eta^2_p = .15$, as shown in Figure 3(a) and (b). Skilled baseball athletes were more accurate than novices only for fastball and curveball discriminations. There were no other skill group-related interactions. However, Pitcher interacted with Pitch-type, $F(2,58) = 10.16$, $p < .01$, $\eta^2_p = .26$, with differences in accuracy between the fastball and curveball only for Pitcher A (see Figure 3(a)), as well as with OP, $F(2,58) = 4.83$, $p = .01$, $\eta^2_p = .14$, with accuracy increasing with more information only for pitcher B (data not shown).

Response sensitivity

In Figure 4, frequency of hits and false-alarms are shown for both pitcher A and B when comparing fastballs to change-ups (collapsed across OP). A hit was a correctly identified

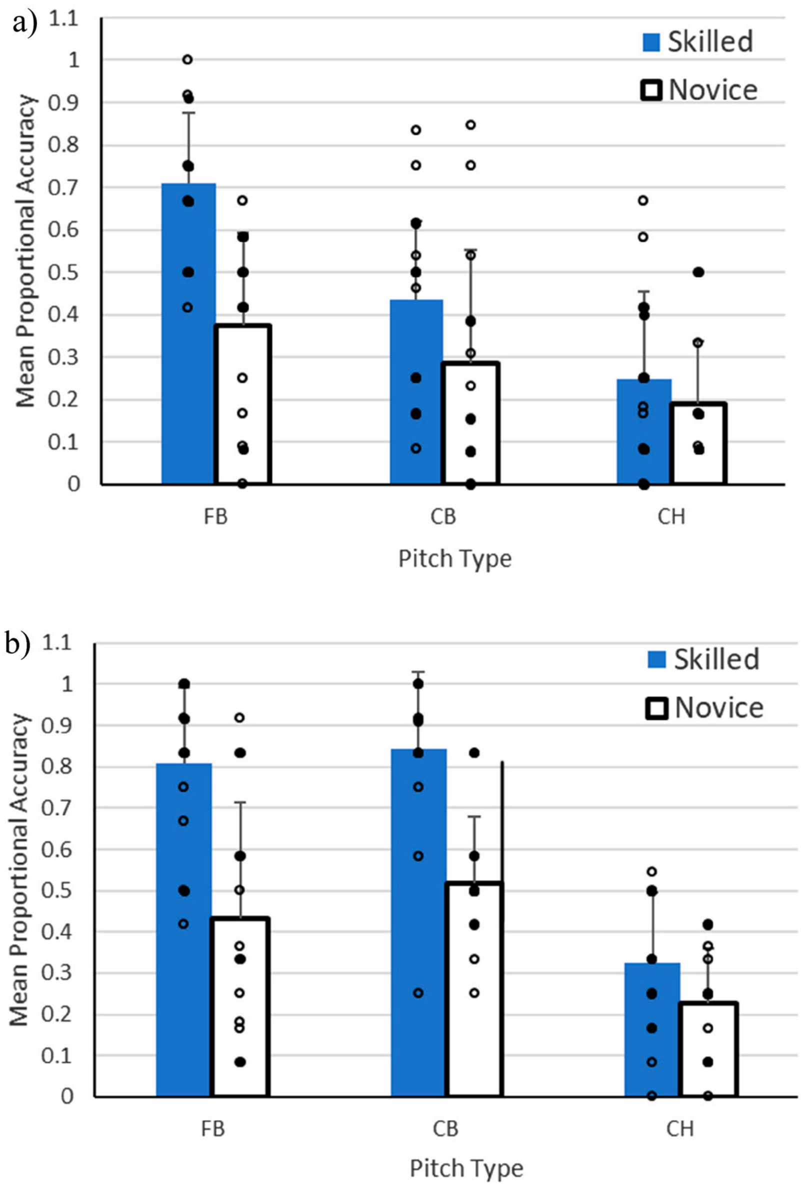

Figure 3. Mean proportional accuracy by pitch type (FB = fastball, CB = curveball, CH = change-up) for pitcher A (a) and pitcher B (b) for skilled and novice participants. Error bars represent SDs and individual mean data points are shown in black. Unfilled dots represent a single data point, filled dots represent overlapping data points.

fastball and a false alarm was a change-up identified as a fastball. Based on a Skill group X Pitcher mixed ANOVA, there was a main effect of group when comparing across the adjusted d' values; with the skilled group ($M_{d'} = 1.18$, $SD = 1.26$) showing higher sensitivity than novices ($M_{d'} = .13$, $SD = 1.01$), $F(1,29) = 12.54$, $p = .001$, $\eta_p^2 = .31$. Congruent with the accuracy data, participants were more sensitive in detecting differences to the pitches thrown by pitcher B (adj $d' = 2.0$) than to those thrown by pitcher A (adj $d' = .36$), F

(1,29) = 8.15, $p = .008$, $\eta_p^2 = .23$. The increased sensitivity was a function of the higher number of false alarms for pitcher A in comparison to B. Close to eighty percent of Pitcher B's fastballs were correctly identified (9.4/12 trials), with false alarms on only 18% of trials.

Response times

In Table 2 we report response times as a function of skill group and condition. The pattern of data mirrored what we had seen for accuracy. Novice participants were slower ($M =$ 838 ms, $SD = 369$ ms), than skilled participants ($M = 625$ ms, $SD = 280$ ms), $F(1,29) = 7.14$, $p = .012$, $\eta_p^2 = .20$ and response times varied as a function of pitcher, $F(1,29) = 4.30$, $p = .047$, $\eta_p^2 = .13$, with participants responding to pitcher A ($M = 769$, $SD = 323$ ms) more slowly than to pitcher B ($M = 672$, $SD = 350$ ms). There was also a main effect for pitch type, $F(2,58) = 3.56$, $p = .035$, $\eta_p^2 = .11$, with fastballs responded to faster overall (as compared to change-ups, $p < .05$), and OP, $F(2,58) = 35.91$, $p < .001$, $\eta_p^2 = .55$, when more information was available, as was the case for OP3 (+366 ms), response times were faster than OP1 (+133 ms) ($p < .05$). Pitch-type and OP also interacted, $F(4,116) = 6.26$, $p < .001$, $\eta_p^2 =$ .18. The difference between OPs was less for change-ups than for the other pitch types (particularly OP1 and OP2). There was also a Pitcher X Pitch type interaction, $F(2,58) =$ 4.16, $p = .02$, $\eta_p^2 = .13$. The slower responses for pitcher A versus pitcher B were mostly driven by slower responses for change-ups.

To determine whether there were speed-accuracy trade-offs for this online discrimination task, we looked at the relationships between accuracy and response time for each pitcher and for each skill group. For skilled participants, there was a small-to-moderate size positive relationship between individual average response times and their

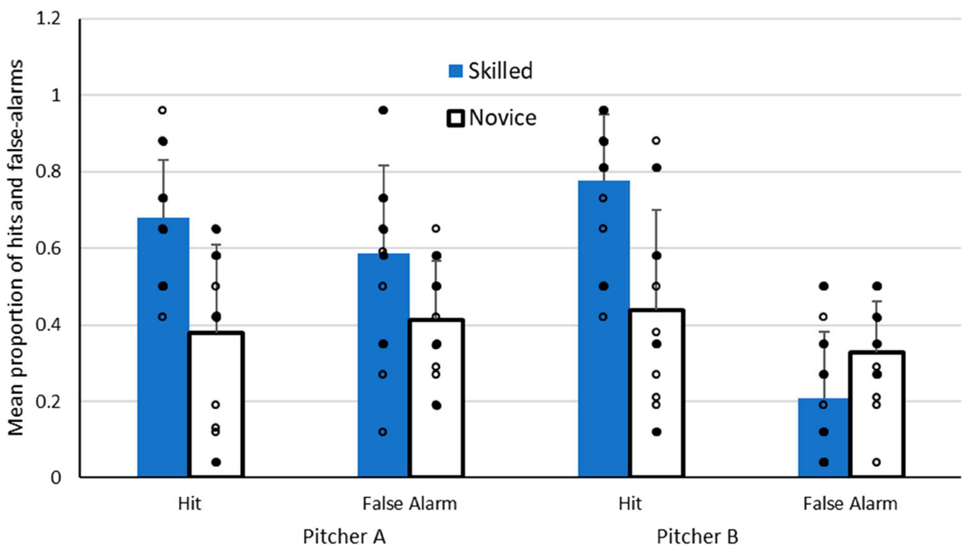

Figure 4. Mean proportion of hits and false alarms in detection of fastballs compared to change-ups as a function of skill group and pitcher. Individual data points are represented by black dots. Unfilled dots represent single data points and filled dots represent overlapping data points. Error bars represent SDs.

Table 2. Mean response times (ms) for both skill groups for pitcher A and B, as a function of pitch type and occlusion point (*OP* = occlusion point).

		Fastball			Change-up			Curve Ball		
		OP1	OP2	OP3	OP1	OP2	OP3	OP1	OP2	OP3
Pitcher A	Skilled	809.82	611.08	543.18	733.76	835.11	663.83	787.73	736.94	651.77
	Novice	894.01	818.65	631.04	946.91	1020.92	801.42	919.52	782.03	780.31
Pitcher B	Skilled	719.00	541.04	469.84	581.26	558.46	409.79	714.56	458.08	418.91
	Novice	956.88	762.15	708.71	877.79	842.61	783.85	956.72	791.39	803.34

overall mean accuracy for pitcher A ($r = .31$, $p = .004$) and pitcher B ($r = .14$, $p = .003$). Generally, slower responses were more accurate, suggestive of a trade-off. However, for novices there were negative relationships between these two dependent measures for pitcher A ($r = -.20$, $p = .008$) and pitcher B ($r = -.32$, $p < .001$). Being slower did not aid accuracy, but rather suggest that being slower was more associated with not knowing the answer.

Within-task reliability

Agreement across the same clips for pitcher A and for pitcher B is shown in Figure 5, as a function of skill group and pitch-type (averaged across OP). Percent agreement values ranged from 50% to 82% for the skilled participants ($M = 71\%$, $SD = 22\%$), with agreement being above 75% for all pitch types for Pitcher B. These values were reduced for novices ($M = 46\%$, $SD = 21\%$), irrespective of pitcher, ranging from 36% to 52%. These descriptive differences were confirmed by a main effect of skill group, $F(1,29) = 27.0$, $p < .001$, $\eta_p^2 = .49$ and pitcher, $F(1,29) = 7.44$, $p = .01$, $\eta_p^2 = .21$, with reliability being significantly less for pitcher A ($M = 54\%$, $SD = 22\%$), than pitcher B ($M = 65\%$, $SD = 26\%$). There was also a main effect of pitch type, $F(2,58) = 3.58$, $p = .03$, $\eta_p^2 = .11$. Post hoc analysis showed that change-ups were responded to with less agreement across trials than curveballs (p

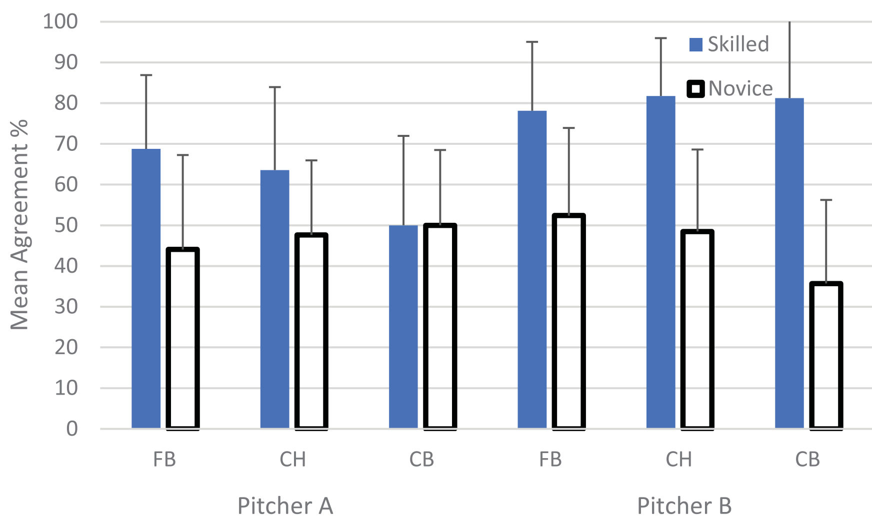

Figure 5. Mean percentage agreement in responses across clips for pitchers A & B as a function of pitch type (FB = fastball, CH = change-up and CB = curveball). Error bars represent SDs.

= .03), but not fastballs ($p = .06$). The only interaction was a 3-way interaction of skill Group with Pitcher and Pitch Type, $F(2,56) = 7.81$, $p = .001$, $\eta_p^2 = .22$. From inspection of Figure 5, differences between the skill groups in terms of % agreement were more pronounced for fastballs for pitcher A, but for curveballs for pitcher B.

Pitch discrimination: in-person and online comparisons

Proportional accuracy

Only accuracy was measured for the online task for a subset of the skilled athletes. As can be seen in Figure 6(a) (in-person) and 6b (online), it was the online task ($M = .62$, $SD = .34$) that was completed more accurately than the in-person task ($M = .52$, $SD = .37$). A 4-way fully RM ANOVA confirmed these differences across tasks, $F(1,8) = 13.23$, $p = .007$, $n_p^2 = .62$. There were again pitcher, $F(1,8) = 34.73$, $p < 0.001$, $n_p^2 = .81$, and pitch type, $F(2,16) = 39.16$, $p < .001$, $n_p^2 = .83$ main effects, but no effect of OP. Pitcher B ($M = .67$, $SD = .34$) was responded to more accurately than pitcher A ($M = .47$, $SD = .35$) and the pitch type effects mirrored those for the online task, with fastballs being responded to more accurately than change-ups only ($p < .01$). We were primarily interested in any task-type interactions that would suggest differences in the pattern of responding when standing ready to bat, versus seated at a computer. However, there were no significant interactions involving Task. All Fs were <1 with the exception of Task X Pitch-type, $F(2,16) = 2.12$, $p = .15$, $n_p^2 = .21$ and Task X Pitch-type X OP, $F(4,32) = 2.33$, $p = .077$, $n_p^2 = .23$. The in-person task (Figure 6(a)), showed a trend for greater differentiation between the three pitches, particularly at the earliest occlusion point (OP1) compared to the online task (Figure 6(b)).

Within-task reliability

There was more agreement in responses for the same clip for the online task ($M = 74\%$, $SD = 18\%$) than the in-person task ($M = 55\%$, $SD = 12\%$), $F(1,8) = 19.76$, $p = .002$, $\eta_p^2 = .71$. Although there were no overall differences as a function of pitcher or pitch type, there was an interaction between Task and Pitcher, $F(1,8) = 13.29$, $p = .007$, $\eta_p^2 = .62$. Post hoc analysis showed that there was significantly less agreement in responses to pitcher B in the in-person task than in the online task ($p = .002$), but not so for Pitcher A.

Between-task reliability

We calculated between task reliability by comparing responses to clips online with responses given in person. Between task agreement was ∼60%. There were no differences in between task agreement as a function of pitcher or OP, only a trend for a pitch type main effect, $F(2,14) = 3.20$, $p = .072$, $\eta_p^2 = .31$ and a Pitcher X Pitch type interaction, $F(2,14) = 8.72$, $p = .003$, $\eta_p^2 = 0.56$. Agreement was generally higher for fastballs (69%) than curveballs (60%) and change-ups (51%), except for Pitcher A, where curveballs showed more between task agreement than change-ups.

Dynamic visual acuity: online task

There were no differences across the skill groups for the online DVA task, $F(1,29) = .74$, $p = .40$, $\eta_p^2 = .025$ (see Figure 7). There was a significant effect of gap size, $F(3,87) = 42.42$, $p < .001$, $\eta_p^2 = .59$. Based on post hoc comparisons, the smallest gap (1 pixel) was responded

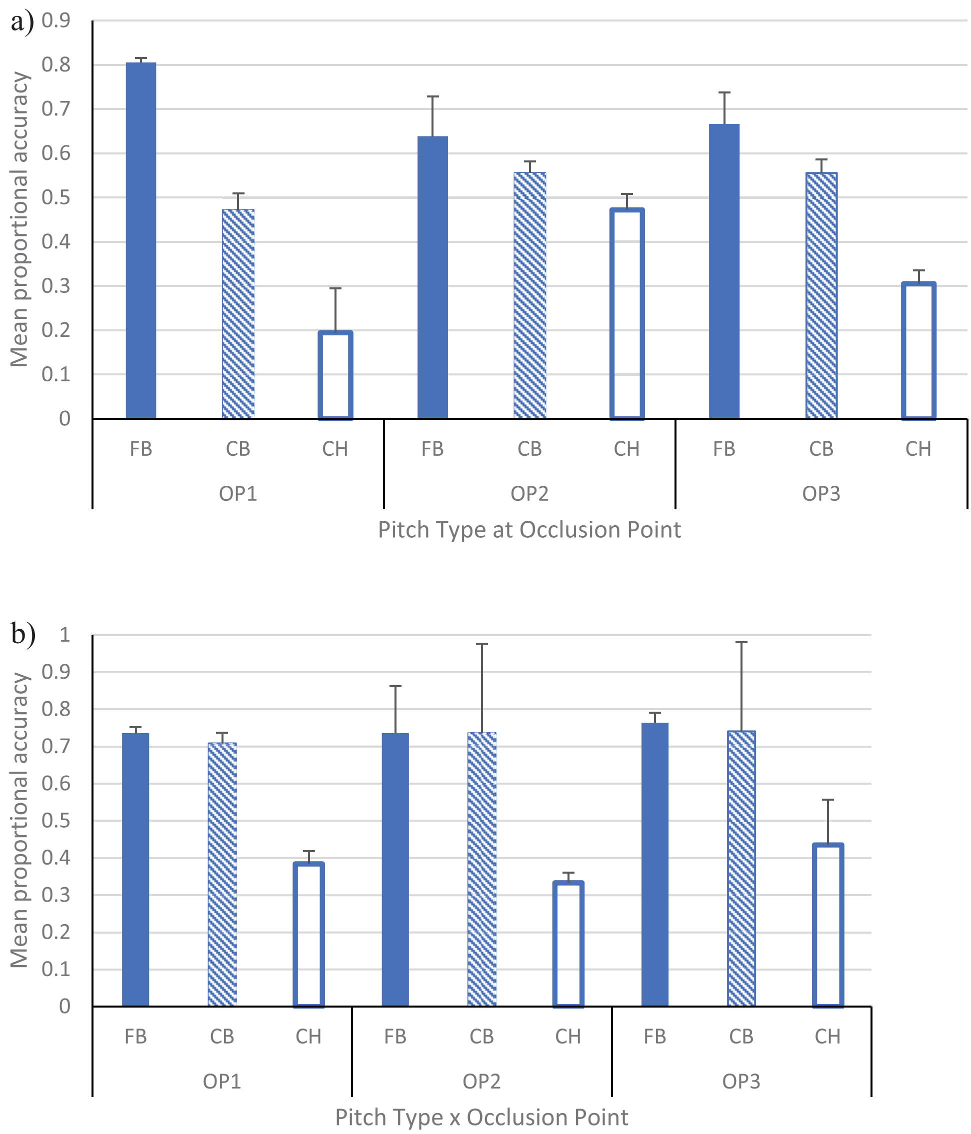

Figure 6. Mean proportional accuracy for the subset of skilled athletes who completed both the in-person (a) and online (b) tasks. Data are shown for each occlusion point, for each pitch type (FB = fastball, CH = change-up and CB = curveball), collapsed across pitcher. Error bars represent SDs.

to significantly less accurately ($p < .001$) than all other gaps (2-4 pixels). There was no interaction between Skill group and Gap size, $F < 1$.

For skilled participants, there was a non-significant trend for a positive relationship between average accuracy on the prediction task and DVA accuracy for pitcher A ($r = .30$, $p = .23$), but a trend for a negative relationship for pitcher B ($r = -.27$, $p = .34$), as illustrated in Figure 8(a) and (b) respectively. For novices, there were positive relationships between prediction task accuracy and DVA accuracy for both pitcher A ($r = .49$, $p = .08$) and pitcher B ($r = .57$, $p = .03$), as shown in Figure 8(c) and (d).

Discussion

In general, we saw high between-group discriminability (validity) and between-pitch sensitivity for the online prediction task, coupled with high reliability for the skilled participants, particularly for Pitcher B video clips. The pattern of results was similar across the online and in-person task, with higher accuracy and greater reliability (based on % agreement) for the online than in-person task. The clips of pitcher A generally resulted in less reliable and less accurate responding than those of pitcher B. There were no skill-group differences in the Dynamic Visual Acuity (DVA) task even though all participants were sensitive to changes in gap size. Although there were relationships between accuracy on the prediction task and the DVA task, it was only the novices who showed a statistically significant positive relationship. In the following paragraphs we expand on these summary points and relate the data to predictions and past literature as well as making recommendations for future work. We first discuss the data with respect to considerations for internal and external validity, before discussions concerning reliability.

Considerations for internal and external validity

In terms of the ability of the pitch discrimination task test to differentiate across skill groups, there was good evidence of discriminability. Congruent with previous studies, novices showed low accuracy on this sport-specific test of pitch-type prediction, performing at chance levels, compared to skilled baseball athletes that were significantly above chance (e.g., Abernethy et al., 2001; Aglioti et al., 2008; Loffing et al., 2015). Although both skill groups were more accurate at the later versus early occlusion points on the online task as would be expected, we did not get any Skill group X Occlusion point interactions indicative of larger differences between the groups for the early occlusion points (i.e., just after ball release). We acknowledge that the two groups in the present study were at near opposite ends of the expertise spectrum, such that it is

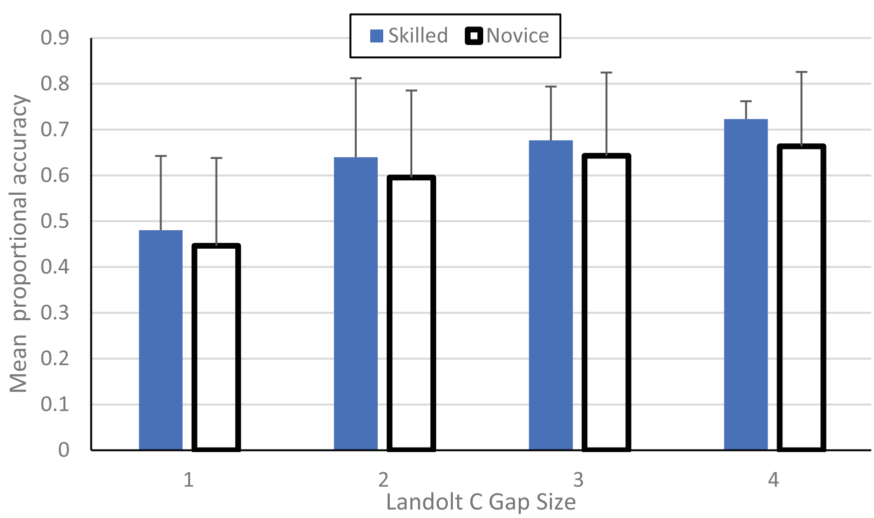

Figure 7. Mean proportional accuracy on the DVA task for the two skill groups at each gap size. Error bars show between participant SDs.

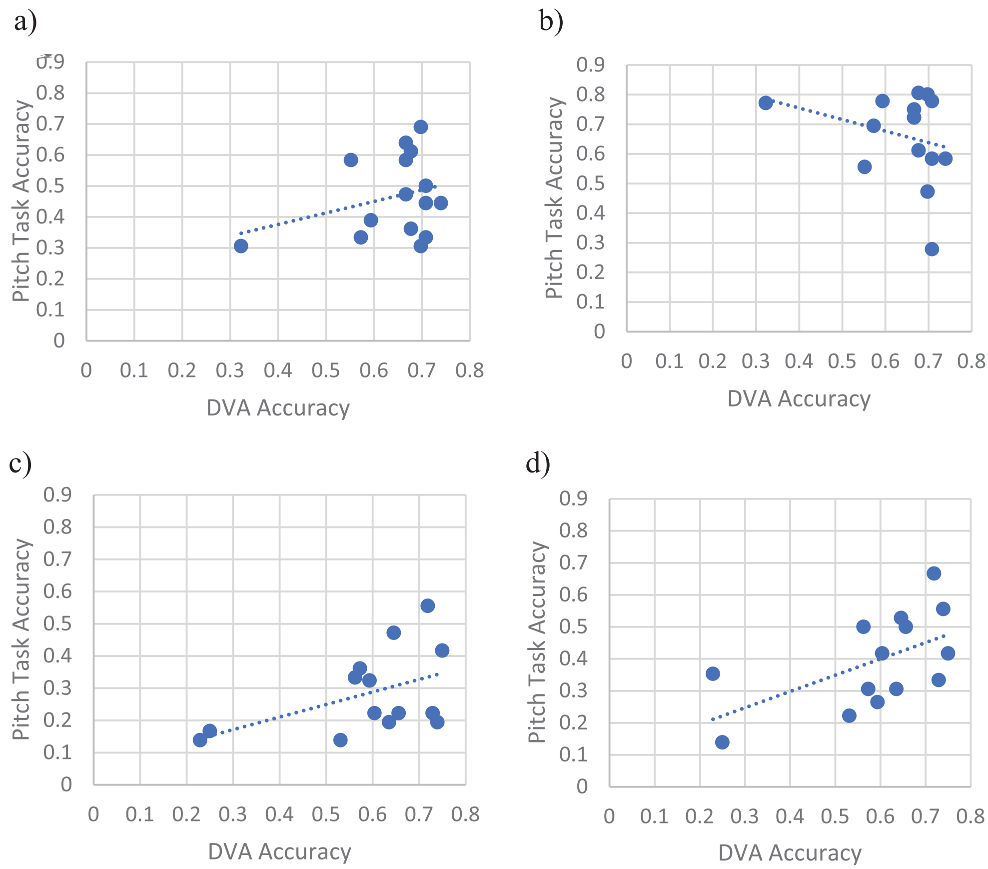

Figure 8. Scatter plots to show the relationship between accuracy on the DVA task and accuracy on the pitch discrimination task for skilled participants for pitcher A (a) and pitcher B (b) and for novice participants for pitcher A (c) and pitcher B (d).

difficult to make any strong statements about the sensitivity of this task to skill. Further testing will be necessary to expand on the validity of this tests and others like it with respect to skill-level sensitivity at between and within-group levels (perhaps with respect to game statistics).

We also showed enhanced perceptual sensitivity in discriminating across pitches that have similar kinematics at the start of the throw (i.e., fastball and change-ups) for the skilled athletes. Not only were skilled baseball athletes able to correctly predict a fastball, something that is thought to be a default response for athletes (Gray, 2010), but they also showed the capability to distinguish a fastball from a change-up, as evidenced through a relatively lower rate of false-alarms to hits, notably for pitcher B (confirmed through analysis of d prime). Novices showed similarities in their responding to hits and false-alarms for both pitchers. Therefore, there is evidence that this online task is able to tap into expertise related to pitch discrimination, given that the skilled athletes were performing significantly above chance, with close to 60% accuracy and high discriminability, especially when differentiating between fastballs and change-ups.

With respect to external validity, we compared data collected from a subsample of skilled participants, who had the opportunity to complete in-person testing in the baseball centre, to their online performance for the pitch discrimination task. We showed a similar pattern of results across both tasks and there were no statistically significant interactions involving task type. However, there were significant differences in accuracy across the tasks, with surprisingly, the in-person task being responded to less accurately than the online task. In the in-person task conducted in the indoor university baseball facility, the main change in procedures was the requirement for participants to hold a bat in a ready-to-swing stance in front of a relatively large standardised screen, with a standard viewing angle and an experimenter present. The addition of these variables led to a significant reduction in accuracy, despite the increased contextual realism from standing, being in a baseball setting and being in a ready-to-swing stance. These variables may have served to have added some evaluative stress or increased task complexity for the individuals not present when doing the task at home alone.

There has been discussion in the literature about whether predictive decisions need to be coupled to the motor response for them to be valid measures of expertise. There is some evidence that for experts, accuracy is higher when the response is realistic and coupled in time to the perceptual stimuli (e.g., Farrow & Abernethy, 2003; Mann et al., 2010; yet see Ranganathan & Carlton, 2007, who showed the opposite). In a recent meta-analysis of cognitive skills in sports, the authors concluded that although sport-specific stimuli were important in discriminating across skill groups through decision making tasks similar to the one we employed here, having a sport-specific response had no added effect on the discriminability of the test (Kalén et al., 2021). Although in neither the in-person or online case were individuals making a real-coupled action response, our results also show that performing the task in an online environment, devoid of baseball related cues, had no negative effects on prediction capability.

Although our pitchers were both highly skilled and representative of pitchers that our participants would likely face in their competitive environment, there could still be questions concerning the generalisability of data based on these two pitchers. The online response accuracy for pitcher B was ~65% for the skilled group, but dropped to about 47% for the same group when responding to pitcher A. Pitcher A had slightly faster pitches (see Table 1), which may have added to the difficulty in discriminating pitch type. However, there is no ideal response accuracy for these perceptual-cognitive tasks, with a need to ensure that they are neither too easy nor too hard, allowing for discriminability based on skill and potential sensitivity to practice-related factors. Moreover, there is reason to think that a success rate of ~50-60% in pitch discrimination may be quite good when considered at high levels of baseball competitive play. Several statistics can be used to assess a player's pitch discrimination in game. For example, slugging percentage is often used as an assessment of power in hitting, addressing how many bases a player records at each at-bat, with the MLB average being .42 in 2020 (Childs, 2001). Assuming that pitch discrimination is a significant factor impacting slugging percentage, as has been argued in prior literature based on biomechanical analysis of weight transfer (e.g., Katsumata, 2007) and correlations between pitch discrimination and in game statistics (Morris-Binelli et al., 2018; Müller & Fadde, 2016), our accuracy scores for both pitchers are in line with actual in-game accuracy.

If we look at the response time data as a function of accuracy for this online discrimination task, for the skilled group there was evidence that slowing down improved accuracy. Baseball batters typically have less than 400 ms to respond to a pitch for a fastball in a Major League Baseball game (e.g., Müller et al., 2017; Gray, 2021). As such, if these participants were to respond to pitcher A under the time constraints of ball travel (whose fastball speed was 89 mph), they would have likely performed poorly. Although we allowed participants up to 3000 ms to respond, the image of the pitcher was no longer available once the clip had finished, so there were constraints on decision time, if not as stringent as would be in place under real-world conditions. More time to respond did not aid accuracy for the novice performers and in fact there was a negative correlation between speed and accuracy. Being slower was more associated with having higher error, suggesting that having additional time to respond did not aid answer generation.

The results of the study were limited by the relatively small sample size during the in-person task. In future studies, it will be important to not only increase sample size, but also to validate individual clips, particularly with different skill groups, and potentially based on hit statistics across a season. Improvements in hitting statistics could be related to improvements on pitch discrimination tests across a larger sample to provide additional tests of validity. Because of COVID-19 restrictions on in-person research activities during the time this experiment was launched, we were limited in collecting more video footage from pitchers. As is clear from the data, there were some issues associated with discriminating pitches for Pitcher A. The data generated from this study therefore reinforce the need to ensure footage used for stimuli is sensitive to skill group differences and reliably responded to over multiple viewings. These assessments and criteria will allow conclusions about the overall validity of 2D video footage (and video training apps) for games and assessments and whether such tests are a valid measure of skill in baseball.

Showing that online pitch discrimination tasks discriminate across athlete skill levels in baseball does not justify whether, or guide how, individuals should use them for practice. Fadde (2006) showed a relationship between in game statistics and use of a video-based pitch discrimination training protocol with high level baseball players. However, there was no evidence that "better" game statistics of the practice group were a direct result of video training, nor that their statistics improved with regular and consistent training because a posttest-only design was used (rather than follow-up retention testing).

Based on past literature (e.g., Uchida et al., 2013; Klemish et al., 2018), there was reason to think that dynamic visual acuity (DVA) would differentiate across skill groups and potentially be related to prediction accuracy if the ability to track moving objects is an important component skill of pitch prediction. We anticipated that a relationship between DVA and pitch discrimination accuracy would speak to the construct validity of our online measure of discrimination. Improved DVA may be a consequence of playing baseball, because the Landolt C task requires similar visual fixations and tracking to those required when watching a baseball (Laby et al., 2019; Uchida et al., 2013). We did not, however, show differences between novices and skilled participants on the DVA task. For the skilled athletes, there were only small correlations between performance on the DVA task and the experimental pitch discrimination task (actually negative for Pitcher

B), whereas it was the novice group that showed positive relationships. These skill-based differences may indicate that skilled participants were able to use early kinematic cues (e.g., grip, shoulder position) rather than tracking ball flight, to better guide their predictions (assuming that ball flight tracking would be a necessary component similar to both the DVA task and the prediction task). If the performance of novices was more dependent on their ability to track a moving object, we would more likely expect a relationship between performance on both tasks. However, novices did relatively poorly on the pitch discrimination task, performing at chance, such that any relation might speak more to an increased motivation to do both tasks well, rather than underlying shared component skills.

Dynamic visual acuity is frequently assessed in studies of athletes involved in high-speed interceptive sports (e.g., Burris et al., 2018; Klemish et al., 2018; Laby et al., 2019). Although it may be somewhat correlated with accuracy in a domain-specific prediction task, our data suggest that it is not a determining factor in pitch discrimination accuracy for skilled, college-level baseball athletes. This result might be seen as contrasting to data from both Laby and colleagues (2019) and Burris and colleagues (2018), who showed that DVA was positively associated with in-game statistics related to pitch discrimination and what is referred to as "plate discipline" (i.e., whether to swing or not). However, the data on DVA as a reliable discriminator of skill is rather mixed, where skill group differences in baseball have both been shown (Uchida et al., 2013) and not been shown (Hoshina et al., 2013). We did not compare DVA task results to in game statistics in this study, nor was our skilled group of the elite level (i.e., professional) as evaluated in some of the aforementioned studies. In future, there may be benefits associated with correlating of in-game statistics (e.g., swing rate, walk rate, on base percentage) to both the DVA task and the pitch discrimination task to further assist in assessing validity.

Within task reliability

Percentage agreement between clips for the online task, indexing inter-item reliability, was relatively high for the skilled performers (71%), particularly for pitcher B. Mean percent agreement on clips of pitcher A were between 50 and 70%, whereas the clips from pitcher B were consistently above 70%. The pitches thrown by pitcher A were faster than those of pitcher B and the lower reliability and lower overall accuracy indicates that the speed and perhaps pitching mechanics caused difficulties for discrimination, impacting accuracy and consistency in responses. Indeed, as shown in Table 1, pitcher B consistently threw 4–5 mph slower than pitcher A, across all pitch types. We had a relatively high level of skill in our study sample, so we were surprised that the first pitcher was challenging for these athletes, but it may be that these types of clips will be more sensitive to training interventions.

As with the accuracy data, it was the online task that showed higher % agreement across clips than the in-person task (74% vs 55% respectively), with greater consistency in responding. The agreement across all clips for both tasks was 61%, which means that there was consistency in the pattern of responding for 22 out of 36 clips for the subsample of athletes that completed both online and in-person tasks. There was a trend for the fastballs to be responded to most reliably, which is congruent with the accuracy data.

Theoretical and applied implications

The factor that has been shown to best discriminate across skill groups, that of sport-specific visual-perceptual information, was included in our online baseball simulation test (Kalén et al., 2021). Skilled, but not novice, participants were able to use this perceptual information to make predictions (above chance), indicating that our task could capture expertise. Further, this indicates that our online task engaged skilled pick-up of advance and object flight perceptual information, which has been reported to map onto lower and upper body action responses in striking sports (e.g., Morris-Binelli & Müller, 2017). The DVA task did not discriminate skill level, which is consistent with a recent meta-analytic study reporting that general visual skills do not make a major contribution to expertise in sports (Kalén et al., 2021). Indeed, when DVA has been manipulated to be below normal levels, it has not impeded interception in other field-based sport tasks (Mann et al., 2010).

The manner of responding in the online environment was qualitatively similar to that for the in-person environment, when players were standing in a batting cage, ready to bat, with moderate reliability across the two settings. However, because of issues with the reliability across pitchers there is a need to further determine what stimuli best capture between skill differences (i.e., filming and testing of additional pitchers and running item analyses on each video clip) and their sensitivity to factors such as age, experience and pitch interception (i.e., at bat statistics) at multiple time points. Additional reliability and validity testing are needed before being able to substantiate claims concerning the efficacy of such online software for assessment and training. Our work here provides a first, necessary step to developing such a tool and providing cautions in the uptake of commercially available software without knowledge of validity and reliability testing with relevance to how the software is to be used.

Summary and conclusions

We developed an online test of pitch discrimination and DVA to evaluate the validity and reliability of a video-based, online pitch discrimination temporal occlusion task among skilled, high-level baseball athletes. This task was developed against the backdrop of the COVID-19 pandemic and a growing market for such online, video-based perceptual training apps in sports. There was evidence for discriminability of the clips when compared to novice performers, as well as external validity when comparing to an in-person version administered in the baseball setting whilst preparing to bat. There was also evidence of within-task reliability, although as with the accuracy data, skilled performers showed better accuracy and consistency for pitcher B (the slower pitcher), rather than pitcher A. Skilled baseball players also showed more sensitivity to pitch type in their predictions, better discriminating change-ups from fastballs. Given large differences between experience and ability of our two different groups, there is a need to investigate an intermediate skill group to add to the validity of this test to discriminate within a narrower range of skill and experience. There were no skill group differences in accuracy for the DVA task and no significant relationship between prediction accuracy and DVA for the skilled group. As such, there was no evidence that tracking ability in general was an underlying component of skill in pitch-type discrimination. For the novice group, DVA was related to performance on the pitch discrimination prediction task, which may be indicative of the novice group's greater reliance on the tracking of ball flight information or perhaps reflective of concentration and desire to do well on the tasks in general.

For future research, there is a need to continue to refine our methods and stimuli based on further testing of varying skill groups and potentially validating with performance against a live-pitcher or with in-game statistics. This question of the validity and reliability of perceptual-skills assessment in sports is becoming increasingly important with the development of online video-based training apps. Before these tasks can be used to train athletes, it is important to identify some constraints on validity and reliability of these apps. One thing is clear, is that both these factors of validity and reliability depend on the stimuli shown (i.e., pitcher). It is likely that a comprehensive test of various pitchers will be needed to strike the balance between discriminatory and reliable stimuli, as well as getting good external validity. We hope to continue to develop this task, based on these data concerning validity and reliability, to test youth athletes across development. Our aim is to help identify the age that perceptual cognitive skills start to distinguish across skill groups, with the idea that this will alert to time-periods when such skills might be best trained. Identifying methods for how to train these skills will be important for specific skill development and adding sport-relevant variety into training, assuming the tasks prove valid.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery grants awarded to NJH [Award # 004067] and to MS [Award # 04987].

Data availability statement

The data that support the findings of this study have been uploaded to the open science framework (Identifier: doi:10.17605/OSF.IO/QRF2W) and are also available from the corresponding author, [NJH], upon reasonable request when it is ethically correct to do so and where this does not violate the protection of human participants, or other valid ethical, privacy or security concerns.

References

Abernethy, B. (1984). Skill in cricket batting: Laboratory and applied evidence. In M. L. Howell, & B. D. Wilson (Eds.), *Proceedings of the VII commonwealth and international conference on sport, physical education, recreation and dance* (pp. 35–50). University of Queensland.

Abernethy, B. (1996). Training the visual-perceptual skills of athletes: Insights from the study of motor expertise. *The American Journal of Sports Medicine*, *24*(6), S89–S92. https://doi.org/10.1177/036354659602406S24

Abernethy, B., Gill, D. P., Parks, S. L., & Packer, S. T. (2001). Expertise and the perception of kinematic and situational probability information. *Perception*, *30*(2), 233–252. https://doi.org/10.1068/p2872

Abernethy, B., Neal, R. J., & Koning, P. (1994). Visual-perceptual and cognitive differences between expert, intermediate, and novice snooker players. *Applied Cognitive Psychology*, *8*(3), 185–211. https://doi.org/10.1002/acp.2350080302

Abernethy, B., & Russell, D. G. (1987). Expert-novice differences in an applied selective attention task. *Journal of Sport Psychology*, *9*(4), 326–345. https://doi.org/10.1123/jsp.9.4.326

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

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*(9), 1109–1116. https://doi.org/10.1038/nn.2182

Aucoin, D. (2019). *Validating Rapsodo 2.0: Pitch design made easier.* Driveline Baseball. https://www.drivelinebaseball.com/2019/07/validating-rapsodo-2-0-pitch-design-made-easier/

Bahill, A. T., & Karnavas, W. J. (1993). The perceptual illusion of baseball's rising fastball and breaking curveball. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 3–14. https://doi.org/10.1037/0096-1523.19.1.3

Brenton, J., Müller, S., & Mansingh, A. (2016). Discrimination of visual anticipation in skilled cricket batsmen. *Journal of Applied Sport Psychology*, *28*(4), 483–488. https://doi.org/10.1080/10413200.2016.1162225

Burris, K., Vittetoe, K., Ramger, B., Suresh, S., Tokdar, S. T., Reiter, J. P., & Appelbaum, L. G. (2018). Sensorimotor abilities predict on-field performance in professional baseball. *Scientific Reports*, *8*(1), 116.

Canãl-Bruland, R., Filius, M. A., & Oudejans, R. R. D. (2015). Sitting on a fastball. *Journal of Motor Behavior*, *47*(4), 267–270. https://doi.org/10.1080/00222895.2014.976167

Chen, Y., Lee, P., Lu, Y., Huang, S., & Yen, N. (2017). Contributions of perceptual and motor experience of an observed action to anticipating its result. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(2), 307–316. https://doi.org/10.1037/xhp0000312

Childs, S. (2001). 2021 *Fantasy baseball: What is average hit rate (AVH)?* Sports Illustrated. https://www.si.com/fantasy/2021/03/05/what-is-average-hit-rate-avh#:~:text = 500%2B%20slugging%20percentage%20is%20good,total%20bases%20divided%20by%20hits

Drew, L. (2021). How athletes hit a fastball. *Nature*, *592*(7852), S4–S6. https://doi.org/10.1038/d41586-021-00816-3

Fadde, P. (2006). Interactive video training of perceptual decision-making in the sport of baseball. *Technology, Instruction, Cognition and Learning*, *4*, 237–256.

Farrow, D., & Abernethy, B. (2003). Do expertise and the degree of perception—action coupling affect natural anticipatory performance? *Perception*, *32*(9), 1127–1139. https://doi.org/10.1068/p3323

Farrow, D., Abernethy, B., & Jackson, R. C. (2005). Probing expert anticipation with the temporal occlusion paradigm: Experimental investigations of some methodological issues. *Motor Control*, *9*(3), 330–349. https://doi.org/10.1123/mcj.9.3.330

Farrow, D., & Reid, M. (2012). The contribution of situational probability information to anticipatory skill. *Journal of Science and Medicine in Sport*, *15*(4), 368–373. https://doi.org/10.1016/j.jsams.2011.12.007

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fooken, J., & Spering, M. (2020). Eye movements as a readout of sensorimotor decision processes. *Journal of Neurophysiology*, *123*(4), 1439–1447. https://doi.org/10.1152/jn.00622.2019

Gray, R. (2010). Expert baseball batters have greater sensitivity in making swing decisions. *Research Quarterly for Exercise and Sport*, *81*(3), 373–378. https://doi.org/10.1080/02701367.2010.10599685

Gray, R. (2017). Transfer of training from virtual to real baseball batting. *Frontiers in Psychology*, *8*(2183).

Gray, R. (2019). Sports training technologies: Achieving and assessing transfer. In N. J. Hodges, & A. M. Williams (Eds.), *Skill acquisition in sport: Research, theory, and practice*, 3rd ed. (pp. 203–219). Routledge.

Gray, R. (2021). Review: Approaches to visual-motor control in baseball batting. *Optometry and Vision Science*, *98*(7), 738–749. https://doi.org/10.1097/OPX.0000000000001719

Gray, R., & Canal-Bruland, R. (2018). Integrating visual trajectory and probabilistic information in baseball batting. *Psychology of Sport and Exercise*, *36*, 123–131. https://doi.org/10.1016/j.psychsport.2018.02.009

Gustafsson, H. C., Young, A. S., Stamos, G., Wilken, S., Brito, N. H., Thomason, M. E., Graham, A., Nigg, J. T., & Sullivan, E. L. (2021). Innovative methods for remote assessment of neurobehavioral development. *Developmental Cognitive Neuroscience*, *52*, 101015. https://doi.org/10.1016/j.dcn.2021.101015

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, and Computers*, *27*, 46–51.

Hodges, N. J., Wyder-Hodge, P. A., Hetherington, S., Baker, J., Besler, Z., & Spering, M. (2021). Topical review: Perceptual-cognitive skills, methods, and skill-based comparisons in interceptive sports. *Optometry and Vision Science*, *98*(7), 681–695. https://doi.org/10.1097/OPX.0000000000001727

Hoshina, K., Tagami, Y., Mimura, O., Edagawa, H., Matsubara, M., & Nakayama, T. (2013). A study of static, kinetic, and dynamic visual acuity in 102 Japanese professional baseball players. *Clinical Ophthalmology*, *7*, 627–632. https://doi.org/10.2147/OPTH.S41047

Hubbard, A. W., & Seng, C. N. (1954). Visual movements of batters. *Research Quarterly for Exercise and Sport*, *25*(1), 42–57.

Kalén, A., Bisagno, E., Musculus, L., Raab, M., Pérez-Ferreirós, A., Williams, A. M., Araújo, D., Lindwall, M., & Ivarsson, A. (2021). The role of domain-specific and domain-general cognitive functions and skills in sports performance: A meta-analysis. *Psychological Bulletin*, *147*(12), 1290–1308. https://doi.org/10.1037/bul0000355

Katsumata, H. (2007). A functional modulation for timing a movement: A coordinative structure in baseball hitting. *Human Movement Science*, *26*(1), 27–47. https://doi.org/10.1016/j.humov.2006.09.005

Klemish, D., Ramger, B., Vittetoe, K., Reiter, J. P., Tokdar, S. T., & Appelbaum, L. G. (2018). Visual abilities distinguish pitchers from hitters in professional baseball. *Journal of Sports Sciences*, *36*(2), 171–179. https://doi.org/10.1080/02640414.2017.1288296

Laby, D. M., Kirschen, D. G., Govindarajulu, U., & DeLand, P. (2019). The effect of visual function on the batting performance of professional baseball players. *Scientific Reports*, *9*(1), 16847. https://doi.org/10.1038/s41598-019-52546-2.

Loffing, F., Stern, R., & Hagemann, N. (2015). Pattern-induced expectation bias in visual anticipation of action outcomes. *Acta Psychologica*, *161*, 45–53. https://doi.org/10.1016/j.actpsy.2015.08.007

Mann, D. L., Abernethy, B., & Farrow, D. (2010a). Action specificity increases anticipatory performance and the expert advantage in natural interceptive tasks. *Acta Psychologica*, *135*(1), 17–23. https://doi.org/10.1016/j.actpsy.2010.04.006

Mann, D. L., Abernethy, B., & Farrow, D. (2010b). The resilience of natural interceptive actions to refractive blur. *Human Movement Science*, *29*(3), 386–400. https://doi.org/10.1016/j.humov.2010.02.007

McLeod, P. (1987). Visual reaction time and high-speed ball games. *Perception*, *16*(1), 49–59. https://doi.org/10.1068/p160049

Milgram, P. (1987). A spectacle-mounted liquid-crystal tachistoscope. *Behavior Research Methods, Instruments, & Computers*, *19*(5), 449–456. https://doi.org/10.3758/BF03205613

Morris-Binelli, K., & Müller, S. (2017). Advancements to the understanding of expert visual anticipation skill in striking sports. *Canadian Journal of Behavioural Science / Revue Canadienne des Sciences du Comportement*, *49*(4), 262–268. https://doi.org/10.1037/cbs0000079

Morris-Binelli, K., Müller, S., & Fadde, P. (2018). Use of pitcher game footage to measure visual anticipation and its relationship to baseball batting statistics. *Journal of Motor Learning and Development*, *6*(2), 197–208. https://doi.org/10.1123/jmld.2017-0015

Müller, S., & Abernethy, B. (2012). Expert anticipatory skill in striking sports: A review and a model. *Research Quarterly for Exercise and Sport*, *83*(2), 175–187.

Müller, S., Abernethy, B., & Farrow, D. (2006). How do world-class cricket batsmen anticipate a bowler's intention? *Quarterly Journal of Experimental Psychology*, *59*(12), 2162–2186. https://doi.org/10.1080/02643290600576595

Müller, S., Abernethy, B., Reece, J., Rose, M., Eid, M., McBean, R., Hart, T., & Abreu, C. (2009). An in-situ examination of the timing of information pick-up for interception by cricket batsmen of different skill levels. *Psychology of Sport and Exercise*, *10*(6), 644–652. https://doi.org/10.1016/j.psychsport.2009.04.002

Müller, S., Dekker, E., Morris-Binelli, K., Piggott, B., Hoyne, G., Christensen, W., Fadde, P., Zaichkowsky, L., Brenton, J., & Hambrick, D. Z. (2023). Attributes of expert anticipation should inform the design of virtual reality simulators to accelerate learning and transfer of skill. *Sports Medicine*, *53*(2), 301–309. https://doi.org/10.1007/s40279-022-01735-7

Müller, S., & Fadde, P. J. (2016). The relationship between visual anticipation and baseball batting game statistics. *Journal of Applied Sport Psychology*, *28*(1), 49–61. https://doi.org/10.1080/10413200.2015.1058867

Müller, S., Fadde, P. J. & Harbaugh, A. G. (2017). Adaptability of expert visual anticipation in baseball batting. Journal of Sports Sciences, 35(17), 1682-1690. https://doi.org/10.1080/02640414.2016.1230225

Palidis, D. J., Wyder-Hodge, P. A., Fooken, J., & Spering, M. (2017). Distinct eye movement patterns enhance dynamic visual acuity. *PLoS ONE*, *12*(2), e0172061. https://doi.org/10.1371/journal.pone.0172061.

Ranganathan, R., & Carlton, L. G. (2007). Perception-action coupling and anticipatory performance in baseball batting. *Journal of Motor Behavior*, *39*(5), 369–380. https://doi.org/10.3200/JMBR.39.5.369-380

Rapsodo. (2019). *Pitching 2.0*. https://rapsodo.com/baseball/pitching/

Scott-Andrews, K. Q., Miller, A. L., Templin, T. J., Hasson, R. E., & Robinson, L. E. (2022). The feasibility and challenges of conducting online research to examine movement behavior in parents and children during the COVID-19 pandemic. *Frontiers in Public Health*, *9*, 720083. https://doi.org/10.3389/fpubh.2021.720083

Smeeton, N. J., Huttermann, S., & Williams, A. M. (2019). Postural cues, biological motion perception, and anticipation in sport. In A. M. Williams, & R. C. Jackson (Eds.), *Anticipation and decision making in sport* (pp. 3–24). Routledge.

Uchida, Y., Kudoh, D., Higuchi, T., Honda, M., & Kanosue, K. (2013). Dynamic visual acuity in baseball players is due to superior tracking abilities. *Medicine & Science in Sports & Exercise*, *45*(2), 319–325. https://doi.org/10.1249/MSS.0b013e31826fec97

Weissensteiner, J. R., Abernethy, B., Farrow, D., & Müller, S. (2008). The development of anticipation: A cross-sectional examination of the practice experiences contributing to skill in cricket batting. *Journal of Sport and Exercise Psychology*, *30*(6), 663–684. https://doi.org/10.1123/jsep.30.6.663

Williams, A. M., Ford, P. R., Eccles, D. W., & Ward, P. (2011). Perceptual-cognitive expertise in sport and its acquisition: Implications for applied cognitive psychology. *Applied Cognitive Psychology*, *25*(3), 432–442. https://doi.org/10.1002/acp.1710