Evaluation of Stereo Acuity in Professional Baseball and LPGA Athletes Compared to Non-Athletes
Melissa Hunfalvay, PhD
Rhonda Orr, PhD
Nicholas P Murray, PhD
Claire-Marie Roberts, PhD

Statement of Conflict of Interest: 2nd, 3rd and 4th authors are independent of the company (RightEye, LLC) with no conflicts of interest. All analysis and results are conducted by the 2nd, 3rd and 4th authors. All data collection was conducted independently of RightEye by experienced vision specialists. The first author works for RightEye as a scientist, is blind to the analysis and follows all ethical considerations during the scientific and experimental process.

ABSTRACT

Purpose: To investigate stereo acuity of professional athletes in an open (baseball) and closed (golf) skilled sport compared to non-athletes using a new digitized version of the Titmus Stereo Test/Wirt Circles. Additionally, to examine the new test for reliability and validity.

Methods: Participants included professional baseball players (PBP; n = 265) and professional women golfers (LPGA; n = 52), and non-athletes (n = 107). All participants were tested on the RightEye Fine Depth Perception (FDP) test and a subset (n = 20) was retested to determine reliability and internal consistency. A different subset (n = 81) was tested for validity comparing the Titmus Stereo Test/Wirt Circles to the RightEye FDP test. Cronbach’s Alpha, intraclass correlation coefficients (ICC) were examined for reliability and ANOVA was conducted to determine convergent validity as well as differences in stereo acuity between groups.

Results: Results: The test was revealed to have high reliability and validity, therefore being a good measure of stereo acuity. Furthermore, significant differences (p<.001) between athletes and non-athletes were found. Both PBP and LPGA athletes were significantly different from non-athletes (p<.05). No differences were found between athlete groups.

Conclusions: The RightEye FDP test is both reliable and valid and can therefore be used confidently as a measure of static stereo acuity. Elite athletes in open and closed skills show significantly better stereo acuity than non-athletes.

Stereopsis is the perception of depth from binocular retinal disparity. Stereo acuity, is one important component of stereopsis. Stereo acuity is the smallest detectable depth difference that can be seen in binocular vision and is critical to human performance as it assists us to see in three dimensions. Stereo acuity is one of the fastest visual depth cues and plays a major role in anticipation timing skills such as catching a ball. Stereo acuity
There may be several reasons why the current body of research seems contradictory. One reason may be the skill levels of athletes for comparison.11,13 Another may be the age of the participants which ranges from children to adults.5 Small sample sizes,12 variation in measurement tools throughout many experiments (e.g. Poltavski & Biberdorf,14 used Nike SPARQ; Paulus et al.,12 used a 3D-TV) and differences between sports (Poltavski & Biberdorf,14 examined ice hockey athletes; Paulus et al.,12 examined soccer players, Laby et al.8 examined baseball players) likely also contribute to the differential findings.

Sports have many different requirements that determine success. One way to differentiate motor skills is determining if they are open or closed skills. Open motor skills are those that involve a nonstable and unpredictable environment where an object or environmental context is in motion and where the performer determines when to begin action.15 Closed motor skills are performed in a stable and predictable environment where the performer determines when to begin the action.15 These motor skill concepts are on a continuum, meaning it is rare to find a skill that is 100% closed or 100% open. However, fielding in baseball is an open skill. Batting in baseball is also open but less open than fielding as the pitcher controls the initiation of the pitch requiring the batter to respond. Golf is a more closed skill than baseball as the golfer can take their time hitting the ball and the ball is not in motion. Stereo acuity is important in both these sports. For baseball, viewing the distance of position players. For golf, precisely knowing the depth the ball is from the athlete is important to the biomechanics of the swing path. It is unclear in the current state of research if athletes at the highest levels, across varying different motor skills, all have better stereo acuity compared to non-athletes. Past research has limitations in terms of great variation in ages within and between groups. Such variation can affect statistical outcomes.16 Development of stereopsis and in-turn stereo acuity is thought
to begin from age 3-5 months,\textsuperscript{17} mature around 7-9 years of age\textsuperscript{18} and deteriorate from age 40 years.\textsuperscript{19}

Variation in tools used to measure stereo acuity could be another contributing factor of equivocal results. The variety of tests and limited information on test validity and reliability bring into question the stimuli used to evaluate stereopsis and therefore the results of the studies.

To date, in the field of optometry the Titmus Stereo Test/Wirt Stereo Test is the most widely used, studied and accepted method of examining stereo acuity, specifically near static stereo acuity.\textsuperscript{20} This test has been determined to be unbiased culturally with high testability for all ages.\textsuperscript{20} The test is based on horizontal retinal disparity of two images and the identification of a shape that is apparently closer. The Titmus test is a two-page design with various stimuli at different angles of disparity. The Wirt circles placed at each corner of a diamond pattern (repeated nine times) are situated on the upper left page (see Figure 1).

The purpose of this study is to examine differences in static stereo acuity using a digitized version of the most widely used static stereo test in optometry, the Titmus Stereo Test/Wirt circles. Furthermore, this study examined the reliability and validity of a new digitized static stereo acuity test: the RightEye Fine Depth Perception test. This study includes adults only whose development of stereo acuity should be at its peak. Additionally, this study compares two groups of professional athletes and non-athletes, and athletes in an open skilled sport (baseball) are compared to athletes in a closed skill sport (golf) for differences in stereo acuity.

\section*{METHODS}

\subsection*{Participants}

52 Ladies Professional Golf Association (LPGA) and 265 professional baseball players (PBP), approximately 40\% playing in the Major League (ML) and 107 non-athletes (NA) were retrospectively obtained for this study through the RightEye database. LPGA participants were between the ages of 18-31 years (M = 24.8, SD = 3.5), professional baseball players were between the ages of 26-31 years (M = 28.1, SD = 3.2) and NA participants were between the ages of 19-35 years (M = 27.1, SD = 5.1), 53\% were male (n = 57) and 47\% were females (n = 50).

To test reliability of the digital translation of the Wirt dots, a small subgroup of twenty participants was re-tested a second time on the RightEye Fine Depth Perception (FDP) test. The subgroup included participants of the NA group between the ages of 22-33 years (M = 26.0, SD = 3.8), 50\% were male (n = 10) and 50\% were females (n = 10).

To test validity the Stereo Fly Test (SFT) specifically the nine Stereotest Circles (SC; see Figure 1) was used on a subgroup of 81 participants. The subgroup included participants of the NA group (n = 40) between the ages of 20-34 years (M = 26.6, SD = 4.8), 45\% were male (n = 20) and 55\% were females (n = 20). The subgroup also included participants of the PBP group (n = 41). All were minor league athletes between the ages of 26-30 years (M = 28.2, SD = 2.8), all were male.
LPGA participants were recruited as part of “The A-Team”, headed by Dr. Don Teig, and included a group of sports trained eye care professionals. PBP participants are RightEye clients and testing was conducted by the sport vision doctor for these teams. NA participants were recruited through advertisements placed on the internet, social media, bulletin boards, and via word of mouth. Testing for the NA group was conducted by sport vision specialists in Maryland. All testers were experienced sport vision specialists (e.g. optometrists, ophthalmologists) and had received and passed the RightEye training, education, and protocol procedures prior to testing.

Participants were excluded from participation in the study if they met any of the following pre-screening conditions: neurological disorders (such as concussion, traumatic brain injury, Parkinson’s Disease, Huntington’s Disease, cerebral palsy); vision related issues that prevented successful calibration of all 9-points (such as extreme tropias, phorias, static visual acuity of greater than 20/40, nystagmus, cataracts or eye lash impediments); small vessel strokes; consumption of drugs or alcohol within 24 hours of testing. All subjects provided informed consent to participate in this study in accordance with IRB procedure (IRB: UMCIRB 13-002660).

Materials and Equipment

The participants were seated in a stationary (non-wheeled) chair that could not be adjusted in height at a desk within a quiet, private testing room. The participants were asked to look at a NVIDIA 24-inch 3D Vision monitor and Alienware gaming system that could be adjusted in height. Participants wore NVIDIA 3D Vision Glasses, and a Logitech (model Y-R0017) wireless keyboard and mouse were used. Participants heads were unconstrained.

Testing Procedure

If participants passed pre-screening requirements and all 9-points of calibration they were then asked to complete the informed consent and demographic information (age, gender, skill level and years playing at a professional level) in the RightEye application. Participants were then seated at 60cm distance from the computer system before testing. They were instructed to don the 3D glasses. Then the fine depth perception test commenced. The participant read the following instructions: You will see 4 circles, press the arrow key on the keyboard that matches the circle that looks like it sticks out. A visual demonstration and two practice trials were given to model appropriate behavior.

To test reliability of the digital translation of the Wirt dots, participants was re-tested a second time on the RightEye Fine Depth Perception (FDP) test. To test validity the Stereo Fly Test (SFT) specifically the nine Stereotest Circles (SC; see Figure 1) was employed. The RightEye Fine Depth Perception (FDP) and the SFT test was conducted immediately after one another. Half the group was randomly selected to be tested on the RightEye FDP test first and the other half on the SFT test first. Instructions outlined in the instruction booklet were followed by each tester. This included holding the booklet straight before the participant to maintain the proper axis of polarization as well as providing good light, but avoiding reflections on the shiny surface. The booklet was held at sixteen inches, measured by each instructor using an extended ruler. The polarized viewers (glasses) were donned for each test.

RightEye Fine Depth Perception Test. The RightEye FDP test is similar in stimuli presentation to the Stereo Fly Test (SFT) specifically the nine Stereotest Circles (SC; see Figure 1). The RightEye FDP test differs from the SC test as the 3D is computer generated and is presented on monitor not in a physical book (see Figure 2).

The RightEye FDP and the SC test use stimuli where one of the four dots are presented with crossed binocular disparity, creating the illusion
that it protrudes (floats) toward the viewer. Target disparities range from 800 to 20 seconds of arc, and decrease as the test progresses. There are 10 levels of disparity in the RightEye FDP 217 test (see Table 1).

Table 1: RightEye FDP levels, arc seconds and centimeters (cm) of disparity

<table>
<thead>
<tr>
<th>Level #</th>
<th>Angle of Disparity (arc seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>800</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>200</td>
</tr>
<tr>
<td>4</td>
<td>140</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>80</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>40</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
</tr>
</tbody>
</table>

One dot at each level is presented in crossed disparity and the location of the disparate dot is randomly selected to reduce learning or memorization effects. Stimuli are shown for 60-seconds or less depending on when the user responds on the keypad. If the user responds in 10 seconds, then the test proceeds to the next stimuli. If a participant fails to respond within 60-seconds that trial is considered a failure.

The administration of the test is conducted in three consecutive phases, each phase has independent logic: phase one is practice, phase two is screening and phase three is testing. In the practice phase a total of two practice trials are given before the screening phase starts. Practice trails are set at 800 and 400 arc seconds. During the screening phase the participant have one attempt per level of disparity to respond. If they respond correctly, then the next smaller level of disparity is shown until participants respond incorrectly. Once they respond incorrectly the testing phase begins at that level. For example, if a person responds correctly in the screening phase at 800 arc seconds, the next trial shows 400 seconds. If participants respond incorrectly in the screening phase at 400 arc seconds, participants begin the testing phase at 400 arc seconds. The purpose of the screening phase is to reduce testing time by getting the participant in the general area of his/her individual disparity threshold before requiring multiple correct responses at the testing phase. Once screening is complete testing begins. In testing, if the subject correctly responds to 3 of 3 or 3 of 4 stimuli then the disparity level gets smaller. The participant sees no more than 4 stimuli at any one level of disparity. If they get 2 incorrect responses the stimuli show a larger disparity. When the participant fails a testing level, that is, the participant gets more than one trial wrong within a level of disparity during the testing phase, the next stimulus is shown at the next highest disparity level, if passed the test ends and the score reported is the last passed level. Output is reported in arc seconds (800-20), levels (1-10) and distance of disparity (2.21-0.06cm).

Validity by Design

Validity by design also considered “face validity” or “priori validity” is concerned with whether the test seems to measure what is being claimed. The RightEye FDP test has several validity by design elements built into the test. These fall into two categories: 1. test stimuli, 2. test logic and flow.

Test stimuli are modeled after a standard clinical measure, the Titmus Stereo Test/Wirt circles. This clinical test is one of the most used tests for static stereo acuity. Relative distances between dots and size of overall stimuli were mathematically calculated to be representative at the 60cm distance required.
for digitized testing. Color, contrast, and arc seconds (distance of disparity) were all mathematically calculated using standard conversion to a 60cm visual angle. Therefore, the RightEye FDP test stimuli were consistent with a known, gold standard, the Titmus Stereo Test/Wirt circles making the translation from booklet to 3D screen comparable and provide validity by design of the stimuli. Test logic and flow: after careful consideration of the Titmus Stereo Test/Wirt circles test and other test logic such as the Amblyopia Treatment Study (ATS) – HOTV Visual Acuity Test Protocol as well as discussions with leaders in the field of optometry and ophthalmology, it was decided that the most effective testing protocol would be the three-phase approach discussed above (see RightEye FDP test). Specifically considering accuracy of results the three-phase logic provided the most likely consideration of: a) limited testing time to reduce the possibility of fatigue; b) reducing the ability to guess due to presenting only one stimuli (25% chance of guessing correctly) at each level of disparity as is the case for the Titmus Stereo Test/Wirt circles test; c) reducing the ability to malinger by controlling time, allowing the participant to have enough time (60-seconds) but not too much time to respond. The randomization of the disparity was also added to reduce a learning or memorization effect which is more likely in the Titmus Stereo Test/Wirt circles test. When a level is incorrectly answered in the Titmus Stereo Test/Wirt circles test the patient is asked to go back to the previous level, which was often answered only a few seconds earlier, and the circle of disparity has not changed. Therefore, unless the patient has a short-term memory loss it is often easily remember as the previous level of disparity is tested within only a few seconds. Error handling, such as known location of the participant’s eyes on the screen further enhances the confidence that the participant was not guessing because the eyes can be confirmed as “on the stimuli” target when the response was made. Furthermore, movement of the participant, left or right, rather than staying still, looking directly at the stimulus, is also measured through error proofing. Should the participant move to one side or the other or move too close to the stimulus this is recorded and will require the participant to retake the test. Furthermore, the Titmus Stereo Test/Wirt circles test booklet may be tipped or moved when held by patients during testing to gain an advantage. All such test logic and flow decisions enhance the RightEye FDP test’s validity by design, providing further confidence in the accuracy of the results.

Data Analysis

Reliability was evaluated using intraclass correlation coefficients (ICC) between trials. In addition, trial-to-trial reliability was evaluated with Cronbach’s Alpha (CA) and the Intraclass Correlation of Coefficients Standard Error of Measurement for each ICC. Alpha level was set at p<0.05 for all statistical test. The ICC indicate the relative reliability and are interpreted using the following criteria ICC > 0.75 specifies excellent reliability and 0.40 < ICC>-.74 represents fair to good reliability.

Convergent validity was investigated by calculating the bivariate correlation coefficient of the Righteye FDP stereo acuity (sec arc) and SC arc seconds. Validity was also examined using the convergent findings within the Analysis of Variance (ANOVA) test.

To test the difference between groups (LPGA, PBP and NA), a one-way ANOVA was used with alpha set at $p = .05$ as the critical level of significance. Tukey’s HSD post hoc test was used when necessary to evaluate significant effects. Also, when necessary, violations of the sphericity assumption were corrected using Greenhouse-Geisser adjustments of the degrees of freedom.

RESULTS

Table 2 shows the means and standard deviations of arc seconds for the three groups. CA, ICC’s, and associated SEM for trial reliability
(Trial 1 & Trial 2) are reported (see Table 2). The CA demonstrated reliability within an acceptable level. Calculated SEMs suggest the measures are capable of accurate assessment of stereo acuity represented in arc seconds. All ICC were statistically significant at the p<0.05 level. The test-retest reliability and internal consistency provides a clear indication these are in fact measuring stereo acuity.

### Table 2: Means followed by standard deviations of groups LPGA, PBP and NA

<table>
<thead>
<tr>
<th></th>
<th>Arc Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPGA</td>
<td>49.42 (25.27)</td>
</tr>
<tr>
<td>PBP</td>
<td>53.35 (18.43)</td>
</tr>
<tr>
<td>NA</td>
<td>94.48 (54.84)</td>
</tr>
</tbody>
</table>

### Table 3: Cronbach's Alpha (CA) and Intraclass Correlation Coefficients (ICC)

<table>
<thead>
<tr>
<th></th>
<th>CA</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arc Seconds</td>
<td>.625</td>
<td>.456*</td>
</tr>
</tbody>
</table>

*p<.5

Convergent validity via a correlation analyses revealed high positive significant correlations between the Righteye FDP test degree of stereo acuity and the SC test (r = .979, n = 82, p < .001). To provide further evidence of validity, the ANOVA for the SC test (F(2, 79) = 6.01, p < .001; ω = .12) produced equivalent group differences with the Righteye FDP test. In addition, validity was also established by design (see Methods section; Validity by Design). A one-way ANOVA was conducted to analyze the differences in stereo acuity between groups. Significant effect of stereo acuity on groups was identified, F(2, 421) = 13.822, p < .001, ω = .45. Tukey’s HSD tests showed that both LPGA golfers and PBP players scored statistically significantly lower (p<.05) than the NA group. However, stereo acuity did not differ between the LPGA and PBP athletes. In addition, the one-way ANOVA for disparity indicated significant effect for groups, F(2, 421) = 10.483, p < .001; 324 ω = .21.

### Discussion

The purpose of this experiment was to examine differences in stereo acuity using a digitized version of the Titmus Stereo Test/Wirt circles. To provide a confidence level in the results the Titmus Stereo Test/Wirt circles were used as they are the most widely used and accepted measure of static stereo acuity. As the test was translated to a digitized version using the ATS-HOTV Visual Acuity Testing Protocol it was important to test the reliability and internal consistency of the digitized version as well as validity. The translation of this standardized test to the 3D digitized version (RightEye FDP test) showed strong reliability. Furthermore, internal test consistency was examined and provided a clear indication that the test measured stereo acuity. Furthermore, validity results between the Titmus Stereo Test/Wirt circles and the RightEye FDP test was high, indicating that the results of the new digitized version can be used and compared accurately to the Titmus Stereo Test/Wirt circles results. These results indicate that both the translation of the visual stimuli and the testing logic (ATS-HOTV) provides similar results to the Titmus Stereo Test/Wirt circles and therefore subsequent analysis between groups can be considered with confidence.

Results between groups revealed that athletes (LPGA and PBP) were significantly better at static stereo acuity measured by the RightEye FDP. However, differences between LPGA and PBP athletes were not significant. These results are consistent with past research by Soloman et al. where professional baseball athletes showed better stereo acuity than inexperienced controls. It is also consistent with past research that shows athletes who play a sport but differ significantly in skill level of that sport also show differences in their stereopsis (Laby et al., baseball; Coffey et al., 348 golf). These results indicate that it does not matter if the sport requires more open skills (baseball) or closed skills (e.g. golf). Superior stereo acuity seems to be important for different sports at the highest level.
Results also reveal less variability in static stereo acuity ability at the highest levels of both golf and baseball. In other words, not only were the athletes significantly better than non-athletes, but the athletes were also more like one another. This further supports that static stereo acuity is significantly better for elite athletes irrespective of gender as the LPGA golfers were all female and the PBP were all male.

Results of this study differ from past research in soccer by Ward & Williams\textsuperscript{13} where no differences were found between static stereo acuity of youth elite and sub-elite soccer players. This may be explained through human development of stereopsis. The participant group in the current experiment was selected based on mature stereopsis and stereo acuity\textsuperscript{18} and before developmental deterioration (after age 40, Lee & Koo\textsuperscript{19}). Ward and Williams\textsuperscript{14} participants included a group of children nine years and younger. Leat et al.\textsuperscript{18} state that stereopsis does not mature in some children until nine years old and that development is not linear in nature. Therefore, if changes in development are still occurring at different rates it could be a confounding variable that accounted for the non-significant differences in the results found by Ward and Williams.\textsuperscript{13}

It is also possible that the nonsignificant results found by Ward and Williams\textsuperscript{13} may occur due to different sport demands between soccer, baseball, and golf. Soccer requires a significantly larger ball than both baseball and golf. Furthermore, the target in soccer (goal area) is many times larger than both the hitting range or catching range in baseball and the cup size in golf. It is therefore possible that soccer has “different enough” demands to both golf and baseball that it could explain why nonsignificant finding were found in soccer, compared to the significant findings found in this study for both golf and baseball compared to non-athletes.

Limitations of this study are considered in terms of broader scope. It would be incorrect to assume a child, who is still developing this ability should be measured against a fully developed person. Also, as children are developing stereo acuity abilities they should be measured against one another within those developmental levels. As development is not linear, that is, some children develop stereo acuity faster than others, longer term predictions about a final level of stereo acuity should be considered with great care. Furthermore, it would incorrect to predict a child’s ability at a sport based on stereo acuity as a potential predictor, before full maturity of stereo acuity is realized. Further research at various developmental levels of stereo acuity need to be examined. Additionally, any prediction of stereo acuity for those who have mature stereo acuity abilities should be considered as only one of many potential factors that may contribute to success in sport. Examination of dynamic stereo acuity as well as distance stereo acuity and stereopsis would provide a more complete picture of the visual abilities of these groups.

CONCLUSIONS

Overall conclusions show that the RightEye FDP test is reliable, internally consistent, and valid and can therefore be used confidently as a measure of static stereo acuity. Elite athletes in open and closed skills show significantly better stereo acuity than non-athletes.

REFERENCES


22. Niehorster DC, Cornelissen THW, Holmqvist K, Hooge ITC, & Hessels RS. What to expect from your remote eye-tracker when participants are unrestrained. Behavior Research. 2017; 1-15. DOI: 10.3758/s13428-017-0863-0


AUTHOR BIOGRAPHY:
Melissa Hunfalvay, PhD
Bethesda, Maryland, USA

A vision scientist and practitioner using eye tracking technology, Dr. Hunfalvay has written, assessed and trained on the science of eye tracking for 20 years. She has worked with professional athletes, military personnel and others for clinical and performance improvement of the visual system.