ResearchOnline@ND

The University of Notre Dame Australia
ResearchOnline@ND

Health Sciences Papers and Journal Articles

School of Health Sciences

2015

Inter-rater reliability and validity of the Australian Football League's kicking and handball tests

Ashley Cripps
University of Notre Dame Australia, ashley.cripps@nd.edu.au

Luke S. Hopper

Christopher Joyce
University of Notre Dame Australia, chris.joyce@nd.edu.au

Follow this and additional works at: https://researchonline.nd.edu.au/health_article

Part of the Life Sciences Commons, and the Medicine and Health Sciences Commons

THE UNIVERSITY OF
NOTRE DAME
AUSTRALIA

Man#3902-2015

Inter-rater reliability and validity of the Australian Football League's Kicking and Handball tests.

Heading title: Australian Football skill tests

Abstract

Talent identification tests used at the Australian Football League's National Draft Combine assess the capacities of athletes to compete at a professional level. Tests created for the National Draft Combine are also commonly used for talent identification and athlete development in development pathways. The skills tests created by the Australian Football League required players to either handball (striking the ball with the hand) or kick to a series of 6 randomly generated targets. Assessors subjectively rate each skill execution giving a 0-5 score for each disposal. This study aimed to investigate the inter-rater reliability and validity of the skills tests at an adolescent sub-elite level. Male Australian footballers were recruited from sub-elite adolescent teams (n=121, age=15.7 ± 0.3 years, height=1.77 ± 0.07 m, mass=69.17 ± 8.08 kg). The coaches (n=7) of each team were also recruited. Inter-rater reliability was assessed using Inter-class correlations (ICC) and Limits of Agreement analysis. Both the kicking (ICC=0.96, P<0.01) and handball tests (ICC=0.89, P<0.01) demonstrated strong reliability and acceptable levels of absolute agreement. Content validity was determined by examining test scores sensitivity to laterality and distance. Concurrent validity was assessed by comparing coaches' perceptions of skill to actual test outcomes. Multivariate analysis of variance (MANOVA) examined the main effect of laterality, with

scores on the dominant hand (P=0.04) and foot (P<0.01) significantly higher compared to the non-dominant side. Follow-up univariate analysis showing significant differences at every distance in the kicking test. A poor correlation was found between coaches' perceptions of skill and testing outcomes. The results of this study demonstrate both skill tests demonstrate acceptable inter-rater reliable. Partial content validity was confirmed for the kicking test, however further research is required to confirm validity of the handball test.

Key Words: Talent identification, skills test, coaches perceptions

Introduction

Australian Football matches are characterised by high running volume and intensities, heavy physical contact and skill executions by both hand and foot (Dawson, Hopkinson, Appleby, Stewart, & Roberts, 2004). The Australian Football League (AFL) coordinates an annual National Draft Combine in order to ascertain if talented athletes have the physical, psychomotor, and psychological capacities required to compete at a professional level (Woods, Raynor, Bruce, McDonald, & Collier, 2015). Since the combine's inception in 1994, physical characteristics of speed, power and aerobic endurance have been examined using a series of physical tests. However, other factors such as technical (Woods, Raynor, Bruce, & McDonald, 2015) skill are likely to impact on performance and selection in Australian Football. Technical skills specific to Australian Football include kicking (the athlete drops the ball from the hands at approximately waist height so that the ball drops towards the kicking foot. Ball-foot contact typically occurs around 0.1-0.3 m from the ground (Ball, 2008)) and handballing (the athlete holds the ball in one hand and strikes the ball, using a clenched fist, with the opposite hand (Parrington, Ball, & MacMahon, 2013)).

In 2009 the AFL introduced a kicking test designed to assess the dominant and non-dominant kicking efficiency of athletes across a range of Australian Football specific distances. In 2010, a handball test was added to the combine test battery which was designed to assess the capacity of athletes to receive the ball cleanly, either on the ground or in the air, and handball efficiently to a target at various distances. Unlike the physical testing measures, such as the vertical jump tests, 20 metre sprint, agility and Multi-Stage Fitness test, which use objective time or distance measures for assessment, the kicking and handball tests are scored subjectively. Assessors subjectively rate skill outcome of both tests using a simple 0-5 Likert scale. However, there are potential limitations when using subjective measures to quantify performance, such as biasing, which may reduce the accuracy or reliability of the skill tests (Thomas, Nelson, & Silverman, 2011). To date, no examination has been conducted to assess the inter-rater reliability of either the AFL's kicking or handball tests.

Physical test results from the AFL combine are used in conjunction with the subjective observations and perceptions of the AFL recruiters, to guide selection in the annual AFL National Draft Combine. Links have been made between physical test performance, professional selection and career success (Burgess, Naughton, & Hopkins, 2012; Pyne, Gardner, Sheehan, & Hopkins, 2005; Robertson, Woods, & Gastin, 2014). Physical tests used have demonstrated both reliability and validity, no such evidence exists for the AFL's skills tests. A simple means of assessing the partial content validity of the kicking and handball tests procedures may be to assess the tests' sensitivity to laterality and distance. Kinematic differences exist between dominant and non-dominant limb kicks (Ball, 2011) and handballs (Parrington, Ball, & MacMahon, 2015) in professional Australian footballers and these

differences are likely to result in accuracy discrepancies. Such dominant and non-dominant limb discrepancies are likely to be further highlighted when the distance to the target increases. Scoring outcomes sensitivity to laterality and distances would indicate partial content validity of the skill tests.

Whilst the skill tests were originally designed for use at the National Draft Combine, they are also commonly used in adolescent development pathways to assess skill efficiency and for talent identification purposes. Test assessors in development pathways are likely to have varying levels of exposure to the test and so scoring variability may occur. Examination of inter-rater reliability using assessors with limited experience scoring the test would provide first evidence that the subjective scoring procedures are reliable when used in this context.

In Australian Football coaches have great insight into an athlete's ability to perform sport specific skills, due to the time spent training and coaching the athletes. As such, examining coaches' perceptions of an athletes' skill may provide a unique means of assessing the concurrent validity of the kicking and handball test procedures. This study aimed to examine the inter-rater reliability, content and concurrent validity of the AFL's skill efficiency tests in adolescent Australian footballers. It was hypothesised that both tests would demonstrate acceptable levels of inter-rater reliability, that laterality and distance would have a significant effect on technical skill outcomes and that coaches' perceptions of skill would correlate with test score outcomes.

Methods

Participants

Male athletes (n= 121, age=15.7 ± 0.3 years, height=1.77 ± 0.07 m, mass= 69.17 ± 8.08 kg) were recruited from seven semi-elite under 16 (U16) Western Australian Football League teams. Athletes and their guardians were given written information sheets detailing the potential risks associated with the study and subsequently provided written informed consent. Coaches (n=7) from each of the teams were also recruited to give a subjective assessment of the skill efficiencies for athletes within their team. The coaches' assessments rated the skills of each athletes in their team on a 1-5 Likert scale. Further detail regarding the coaches' perceptions of skill is provided later. Assessors for the test were all university students with varying levels of exposure to Australian Football. Assessors were given a briefing on the tests purpose and scoring criterion prior to commencement. To further familiarise the assessor with the test they were also required to watch the test conducted at least once prior to being allowed to score the test. Ethics approval was granted by the University's Human Research Ethics Committee.

Procedures

The test procedures for both skill tests are provided by the AFL (Sheehan, 2010). Figure 1 illustrates the layout of the kicking test. Athletes were required to perform three right and three left footed kicks. Athletes ran towards the feeder and received the ball around chest height on the kick line. At the same time as receiving the ball, the feeder instructed the participant to kick to one of six randomly assigned targets. The player then circled the turn cone and kicked to the appropriate target (the targets are other players at the designated

points). The first (20 m) target was set on a 45° angle from the intersect of the kick lines in Figure 1; the second (30 m) and third (40 m) targets were then set directly back from the first target. The target circles were four metres in diameter. Once the kick was delivered, the player returned to the starting point and repeated until all six targets had been called.

Figure 1 near here

Two student assessors stood approximately 35 m from the kick line in order to best assess the kicks. The assessors stood two metres apart aside the designated scoring position and were instructed not to communicate results to each other. Assessors were instructed to judge the kick on the criteria outlined in Table 1.

One point was removed from the possible five points for each kick if; the kick execution took longer than 3 seconds (monitored by the assessors using a stop watch from time of hearing the call from the feeder to skill execution), the kick was executed beyond the kick line, or the kick was executed incorrectly (unconventional flight and or spin). If the participant kicked to the wrong target, a score of zero was given.

Table 1 near here

The handball test is depicted in Figure 2. Athletes received the ball six times and completed six handballs. The athlete received the first three balls from the ground and the second three were thrown to the receiver around chest height. The athlete was required to perform three right and three left-handed handballs. Athletes ran towards the feeder and received the ball on the pick-up line. At the same time as receiving the ball, the feeder instructed the participant to handball to one of six randomly selected target players standing in designated positions. The first (6 m) target was set on a 45° angle from the release line;

the second (8 m) and third (10 m) targets were then set straight back from the first target. The participant was required to handball to the appropriate target, before the release line. Once the handball was delivered, the player jogged around the turn cone and returned to the start point and repeated until all six targets had been called.

Figure 2 near here

Two student assessors stood 5 m behind the feeder to assess the handballs. The assessors stood two metres apart aside the designated scoring position and were instructed not to communicate results to each other. Assessors were instructed to judge the take and handball based on the criteria outlined in Table 1.

One point was subtracted if; the ball gather and handball took longer than 3 seconds to be executed (monitored by the assessors using a stop watch from time of hearing the call from the feeder to skill execution), or the handball was completed beyond the release line. The delivery was given a score of zero if the participant handballed to the wrong target.

Coaches perceptions of the athletes

Prior to receiving the results of the tests, the athletes' coaches were asked to rate athletes from their team on a 1-5 Likert Scale for kicking and handball efficiency, and clean hands (their ability to take the ball cleanly either in the air or on the ground) with rating listed as; 5 rare, 4 excellent, 3 good, 2 marginal and 1 poor in accordance with the AFL youth coaching manual (2004). Outcome descriptors were attached to the 1-5 rating scale. For example, when assessing kicking and handball ability: a 5 mark was given if the athlete was considered very accurate on both dominant, and non-dominant sides, and when under

pressure; the athlete was also required to be a very good decision maker. Coaches were also asked to categorise athletes as right (n=102) or left (n=19) side dominant. If they were unsure they were instructed to leave the field blank. These athletes (n=8) were then excluded from the analysis.

Data Analysis

The kicking and handball tests were assessed for inter-rater reliability, content and concurrent validity. Inter-rater reliability was examined using the subjective scores provided by two independent assessors, who both rated every disposal using the scoring procedure developed by the AFL.

Content validity was assessed by examining the scoring outcomes sensitivity to laterality across a range of Australian Football specific distances. Concurrent validity was assessed by comparing the scores from both tests to coaches' perception of skill efficiency. For the kicking test, the coaches' perceptions of kicking ability was directly compared to their testing score. For the handball test, because the test examines both the ability to receive the ball cleanly and handball efficiently, the coaches' perception of both clean hands and handball efficiency was summated and compared to the testing outcome.

Statistical analysis

Statistical analyses were carried out using SPSS software (Version 22.0, SPSS Inc., USA). Inter-rater reliability was assessed as relative and absolute measures. Relative reliability was calculated by comparing the total score given by both assessors using intra-

class correlation coefficients (ICC). Absolute reliability was calculated using the 95% limits of agreement (LOA) method developed by Bland and Altman (1986).

Scores were reported as means and standard deviations. Multivariate analysis (MANOVA) was used to examine the main effect of "laterality" (two levels: dominant and non-dominant) on the skills test variables. Cohen's d effect sizes (ES) were calculated, with an ES of 0.20 considered small, 0. 50 medium, and 0.80 large (Cohen, 1998). The correlation between actual testing outcomes and coaches' perceptions of skill was assessed using Pearson's correlation coefficients (r). Significance was set at $P<0.05$.

Results

Inter-rater reliability for both the kicking (ICC=0.96, $P<0.01$) and handball tests (ICC=0.89, $P<0.01$) were strong and within the limits of agreement demonstrating acceptable levels of absolute reliability (Figure 3).

Figure 3 near here

The Pillai's trace (V) revealed a significant effect of laterality on the kicking (V= 0.10, $F_{(3, 252)}$ =9.63, $P<0.01$) and handball (V= 0.06, $F_{(3, 252)}$ =2.85, $P=0.04$) tests. Follow-up univariate analysis revealed dominant leg kicks scored significant higher for all distances ($P<0.01$) with medium effects demonstrated. Dominant hand disposals in the handball test only significantly outscored the non-dominant on the long target (ES= 0.30, $P<0.01$) with small to medium effects demonstrated. Short (ES= 0.26, $P=0.09$) and medium (ES= 0.21, $P=0.16$) handballs showed non-significant differences between dominant and non-dominant limbs. Summary of the tests results can be seen in Table 2. There was no significant

correlation between coaches' perceptions of skill and kicking (r=-0.13, P=0.75) or handball (r=0.04, P=0.63) test scores.

A number of delivery errors were made in both tests by the athletes, whereby the athlete passed to the wrong target. A total of 25 errors made in the kicking test (3.23%) and 95 made in the handball test (12.27%).

Table 2 near here

Discussion

Inter-rater reliability

Relative and absolute inter-rater reliability for both the kicking and handball tests was shown to be strong. The results of this study therefore suggest that the use of inexperienced assessors to administer the AFL's skills tests will not affect the reliability of the tests scoring outcomes. Further, considering the assessors came from a varied and somewhat inexperienced football background, it is reasonable to assume that employing assessors with greater assessment experience, such as those used at the National Draft Combine, would further improve the reliability outcomes of the tests. There were a high number of delivery errors in the handball test. The higher number of errors in the handball test may have slightly elevated the test's reliability measures, as it removed the opportunity for scoring variability. However given the strength of the findings in the reliability analysis, these effects are likely to be minimal.

Validity of AFL skills tests

The results of this study demonstrates mixed results when assessing content validity. Scoring outcomes for the kicking test shows a significant ability to differentiate between accuracy on dominant and non-dominant foot kicks, across varying Australian football specific distances. While the handball test was only able to significantly differentiate between laterality, with inconsistent results apparent when examining effects of distance.

As with most skill tests, the AFL's skills tests are closed-skill tests and are unable to examine every component of the complex task assessed (Robertson, Burnett, Cochrane, 2014). Coaches or scientists designing skill tests are therefore required to select the components of a specific skill they wish to examine, with the intended use of the protocols and results in mind. The two AFL skill tests are designed to be used for both elite and sub-elite talent identification and to provide feedback to athletes for development purposes. Specifically the skills test seek to assess the athlete's capacity to accurately dispose of the ball on their dominant and non-dominant limbs, across varying Australian Football specific distances. Therefore the kicking test in this context demonstrates partial content validity, as the scoring outcomes can differentiate between both laterality and target distance. The AFL's kicking test provides an appropriate means of assessing and providing feedback to development athletes regarding their kicking skills. However, further research is required to determine if the kicking test can differentiate between athletes of higher and lower playing abilities or if kicking test outcomes change with age.

The AFL's handball test did not show the same level of content validity demonstrated by the kicking test. Whilst the test was able to differentiate between dominant and non-dominant disposals, it failed to consistently differentiate between target distances. This may be due to the short (6 m) and medium (8 m) distances not being long enough or the task itself being too simple to elicit meaningful accuracy changes. Further research is needed to confirm the use of the handball test for providing a valid means of handball skill assessment.

Both the kicking and handball tests demonstrated poor concurrent validity, suggesting the AFL skills tests results are not representative of coaches' perceptions of athletes kicking and handball skills. The poor concurrent validity of the skill tests is likely due to the tests inability to replicate all match related skill demands. In matches other factors are likely to influence an athlete's skill efficiency by both hand and foot include, opposition pressure, decision making, and fatigue. The poor concurrent validity demonstrated by both tests suggests that coaches should be cautious when using test results to predict match related skill outcomes.

An identified weakness of the handball test is that the test examines two independent skill outcomes but only reports a single score. This means when examining the scoring outcomes it is impossible to tell which of the two skills in the test the player may have excelled or scored poorly in. For example, a player may have fumbled the ball, but executed an excellent handpass; or taken the ball cleanly but executed a poor handpass. In both cases the scoring outcome would not identify which skill the player performed well in and which they did not. A simple suggestion to eliminate this issue is to incorporate two scoring protocols, one for the clean-hands component of the test and a second for the disposal

outcome. A further suggestion to reduce delivery errors in the test may be to adopt a pre-determined delivery pattern. This may reduce any errors associated with the athlete miss-hearing calls or decision making errors.

A limitation of this study was that the kicking and handball tests were originally designed to be used at the AFL National Draft Combine with athletes of eligible draft age (at least 18 years of age before 31st December of the relevant selection year). Whereas, the athletes we recruited were around two years younger than the athletes who would typically perform the test. Further assessments of the tests validity should therefore be conducted with athletes of eligible draft age.

Conclusion

Both the AFL's kicking and handball tests demonstrated acceptable levels of relative and absolute inter-rater reliability. The kicking tests was also shown to demonstrate partial content validity, with the tests able to discriminate between dominant and non-dominant disposals, across a range of Australian Football specific distances. The AFL's handball test was also able to discriminate between laterality, however it could not consistently discriminate between disposal distances. Both tests demonstrate poor concurrent validity, when compared to coaches' perceptions of skill. The AFL's kicking test may provide an appropriate means of assessing and providing feedback to development athletes regarding their kicking skills, with further research required to establish if the handball test is appropriate to do the same. Future research should establish if both tests can differentiate

between athletes of higher or lower playing abilities and if performance in the skill tests improve with age.

Acknowledgements

Key points

- The skill tests created by the AFL demonstrated acceptable levels of relative and absolute inter-rater reliability.

- Both the AFL's skills tests are able to differentiate between athletes dominant and non-dominant limbs. However, only the kicking test could consistently differentiated between score outcomes over a range of Australian Football specific disposal distances.

- Both tests demonstrated poor concurrent validity, with no correlation found between coaches' perceptions of technical skills and actual skill outcomes measured.

References

Australian Football League (2004), *AFL Youth Coaching Manual*: The Australian Football Leauge.

Ball, K. (2008), Biomechanical considerations of distance kicking in Australian Rules football. *Sports Biomechanics* **7(1),** 10-23.

Ball, K. (2011), Kinematic comparison of the preferred and non-preferred foot punt kick. *Journal of Sports Sciences* **29(14),** 1545-1552.

Bland, M. J., and Altman, D. G. (1986), Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* **327(8476),** 307-310.

Burgess, D., Naughton, G., and Hopkins, W. (2012), Draft-camp predictors of subsequent career success in the Australian Football League. *Journal of Science and Medicine in Sport* **15(6),** 561-567.

Cohen, J. (1998), *Behavioural sciences*. Hillsdae, NJ: Lawrence Erlbaum Associates.

Dawson, B., Hopkinson, R., Appleby, B., Stewart, G., and Roberts, C. (2004), Player movement patterns and game activities in the Australian Football League. *Journal of Science and Medicine in Sport* **7(3),** 278-291.

Parrington, L., Ball, K., and MacMahon, C. (2013), Game-based analysis of handballing in Australian Football. *International Journal of Performance Analysis in Sport* **13(3),** 759-772.

Parrington, L., Ball, K., and MacMahon, C. (2013), Kinematics of perferred and non-perferred handballing in Australian Football. *Journal of Sport Sciences* **33(1),** 20-28.

Pyne, D. B., Gardner, A. S., Sheehan, K., and Hopkins, W. G. (2005), Fitness testing and career progression in AFL football. *Journal of Science and Medicine in Sport* **8(3),** 321-332.

Robertson, S., Woods, C., and Gastin, P. (2014), Predicting higher selection in elite junior Australian Rules football: The influence of physical performance and anthropometric attributes. *Journal of Science and Medicine in Sport* **18(9)**, 225-229

Robertson, S., Burnett, A., and Cochrane, J. (2014), Tests examining skill outcomes in sport: A systematic review of measurement properties and feasibility. *Sports Medicine* **44(4),** 501-218.

Sheehan, K. (2010), *NAB AFL National Draft Combine 2010: Testing protocols*. Australian Football League: Melbourne.

Thomas, J. R., Nelson, J. K., and Silverman, S. J. (2011), *Research Methods in Physical Activity.* Champaign, IL: Human Kinetics.

Woods, C. T., Raynor, A. J., Bruce, L., McDonald, Z., and Collier, N. (2015), Predicting playing status in junior Australian Football using physical and anthropometric parametres. *Journal of Science and Medicine in Sport* *18***(2),** 225-229.

Woods, C. T., Raynor, A. J., Bruce, L., and McDonald, Z., (2015), The use of skills tests to predict status in junior Australian football. *Journal of Sport Sciences* *33***(11),** 1132-1140.
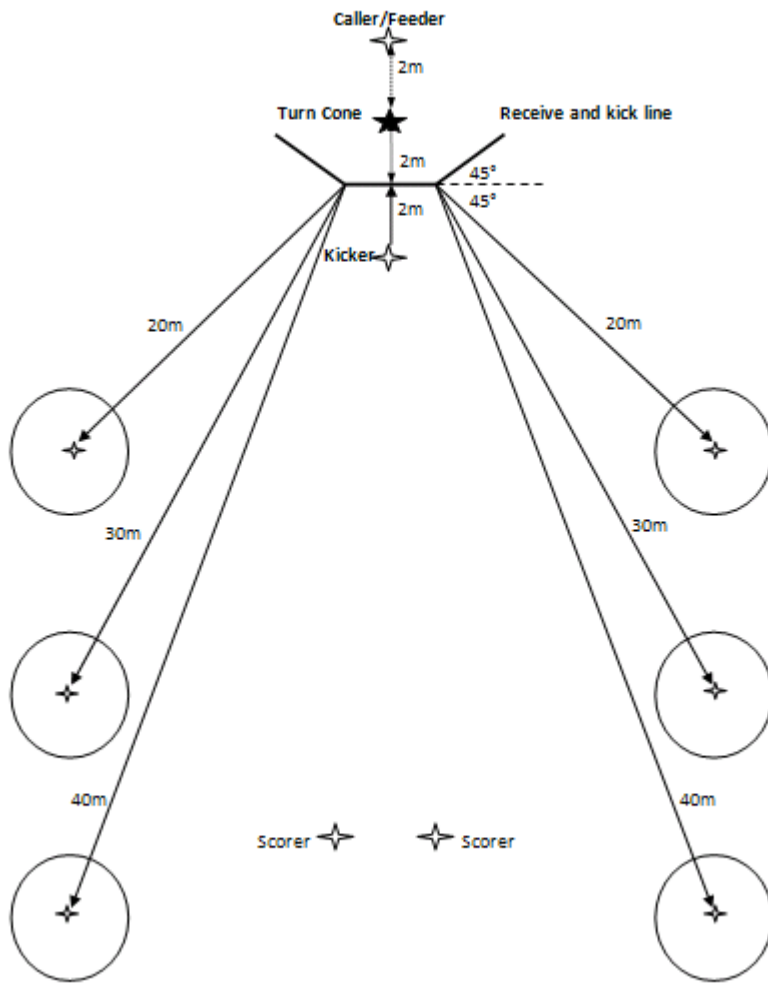
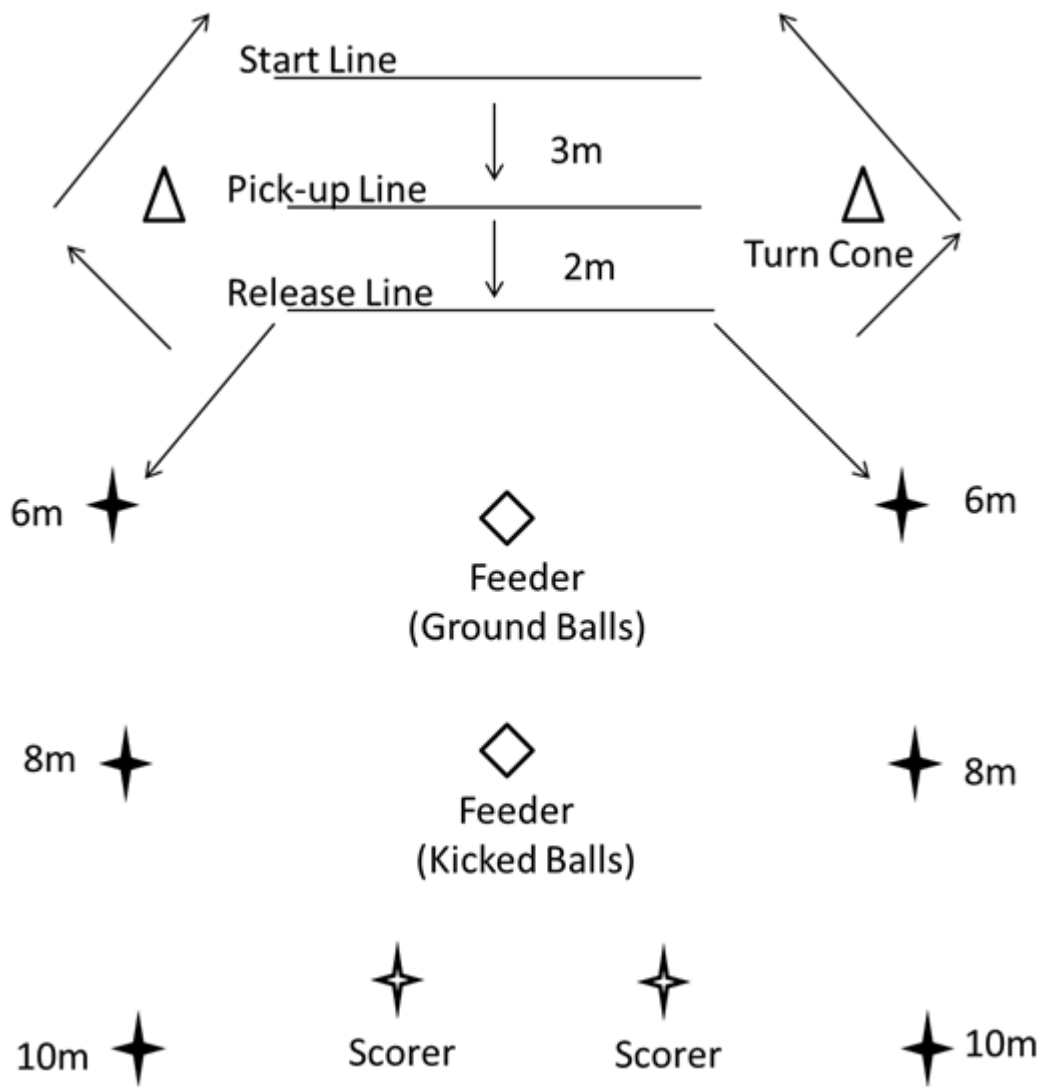Figure 1. Layout of the Kicking Test, adapted from Sheehan (2010).

Figure 2. Layout of the Handball Test, adapted from Sheehan (2010).
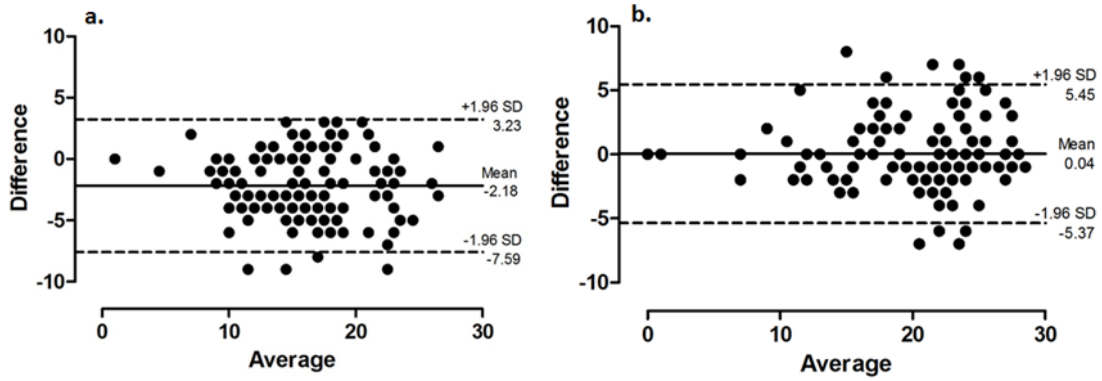
Figure 3. Bland-Altman Limits of Agreement analysis for the Kicking (a) and Handball (b)

Tests

Table 1. Scoring protocols for the AFL's kicking and handball tests.

Points	Rating	Kicking Test	Handball Test
5	Excellent	Target didn't move & ball travelled quickly with low trajectory & perfect spin.	Clean take, quick execution with perfect spin & target not moving receiving ball at chest height
4	Very Good	Target receives within one step of the cone, low trajectory & good spin.	Clean take, quick execution and good spin with target moving slightly to receive
3	Effective	Target receives with a foot inside circle, good trajectory & spin.	Clean take, satisfactory execution with target able to take the ball after moving
2	Ineffective	Target had to leave circle to mark ball, good trajectory & spin.	Fumble but recovers to reach target with good technique
1	Poor	Target unable to mark football, poor trajectory & spin.	Fumbles and gets ball to target with poor technique
0	Fail	Misses target or delivers to wrong target.	Fumbles and misses the target completely

Table 2. Scoring outcomes (mean and standard deviation) for dominant and non-dominant limb disposals for the kicking and handball tests.

Disposal Distance	Kicking Test			Handball Test		
	Dominant	Non-Dominant	Cohen's d	Dominant	Non-Dominant	Cohen's d
Short	3.21 ± 1.39	2.43 ± 1.47*	0.55	3.77 ± 1.50	3.38 ± 1.52	0.26
Medium	2.74 ± 1.39	2.17 ± 1.29*	0.43	3.34 ± 1.46	3.02 ± 1.65	0.21
Long	2.40 ± 1.26	1.90 ± 1.07*	0.43	3.54 ± 1.54	3.09 ± 1.50*	0.30

* Significantly ($P<0.01$) less than dominant limb score