University of South Australia

**Inter- and intra-rater reliability of the Athletic Ability Assessment in sub-elite Australian Rules football players.**

This is the peer reviewed author accepted manuscript (post print) version of a published work that appeared in final form in:

Rogers, Daniel K, McKeown, Ian, Parfitt, Gaynor, Burgess, Darren & Eston, Roger G 2019 'Inter- and intra-rater reliability of the Athletic Ability Assessment in sub-elite Australian Rules football players.' Journal of strength and conditioning research, vol. 33, no. 1, pp. 125-138

This un-copyedited output may not exactly replicate the final published authoritative version for which the publisher owns copyright. It is not the copy of record. This output may be used for non-commercial purposes.

The final definitive published version (version of record) is available at:
https://doi.org/10.1519/JSC.0000000000002175

Persistent link to the Research Outputs Repository record:
http://researchoutputs.unisa.edu.au/11541.2/128732

General Rights: Copyright and moral rights for the publications made accessible in the Research Outputs Repository are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the persistent link identifying the publication in the Research Outputs Repository

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

https://search.ror.unisa.edu.au

**Journal of Strength and Conditioning Research Publish Ahead of Print**
**DOI: 10.1519/JSC.0000000000002175**



# Inter- and intra-rater reliability of the Athletic Ability Assessment in sub-elite Australian Rules football players.

Running Head: Reliability of the Athletic Ability Assessment in sub-elite Australian rules football

Research was conducted at the University of South Australia and Port Adelaide Football Club

Daniel K. Rogers[1,2], Ian McKeown[2], Gaynor Parfitt[1], Darren Burgess[1,2], Roger G. Eston[1]

1. School of Health Sciences, University of South Australia
2. Port Adelaide Football Club

Author Contact: Daniel Rogers, 10 Apollo Street Woodcroft, South Australia, 5162

Email: drogers@pafc.com.au

Mb: 0411 327 645

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

**Abstract**

The aim of this study was to determine the inter- and intra-rater rater reliability of the Athletic Ability Assessment (AAA) in sub-elite Australian Rules football (ARF) players. Eighteen male ARF players completed the AAA movement assessment (overhead squat, double lunge [left and right], single leg Romanian deadlift [left and right], chin-up and push-up), on two occasions separated by one week. During the first movement assessment players were filmed in the frontal and sagittal planes. Ten raters took part in the study (one experienced rater and nine novices) and were assigned in a quasi-random manner, to complete either (a) real-time assessment on two occasions, (b) real-time assessment on one occasion or (c) video-based assessment on two occasions. When assessed in real-time, of the 7 component movements of the AAA, raters registered moderate or greater intra-rater agreement on between 2 and 5 occasions. Intraclass correlation coefficients (ICC) of between 0.50 and 0.61 for the AAA total score indicated poor real-time intra-rater reliability for this variable. When assessed by video-recording, raters registered moderate or greater intra-rater agreement on between 6 and 7 occasions. The ICC for total score ranged between 0.60 and 0.93. Overall poor inter-rater reliability was evident for AAA component movements regardless of whether it was assessed in real-time or from video. Findings suggest the AAA is most reliably employed when assessed via video. It is recommended that if comparison between multiple raters is desired, a stringent training process be applied so that the interpretation of AAA scoring criteria is standardised across raters.

**Key words:** *Movement Competency, Athletic Development, Strength and Conditioning, Testing*

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

## INTRODUCTION

Athletes participating in field and court-based sports require the development of a wide range of physical qualities to optimise athletic performance (17). As such, the development of muscular strength, power and speed have become a crucial part of the training process for many sports (27). However, training methods used to develop these physical qualities demand pre-existing competency in rudimentary athletic movement patterns such as hip-hinging, squatting and lunging if they are to be safely and effectively applied (14,17). In practice, if an athlete does not possess the basic athletic movement competency (AMC) required to undertake a specific movement or training exercise, some areas of training must be regressed to first master more rudimentary patterns. This can hinder athletic development by limiting the diversity of training methods to which an athlete can be safely exposed. It therefore seems pertinent to develop and maintain AMC at a standard which allows for the effective application of developmentally appropriate training methods, aimed at maximising athletic development.

Athletic movement competency is typically quantified via the application of movement screens or assessments – processes which involve the rating of a combination of movements based on subjective criteria (6, 16). Each component movement is individually scored and combined to provide a total score reflecting overall competency in the battery of movements performed (6, 16). If the goal of a training program is to improve AMC, the reliability of this process becomes critical as practitioners must be confident of the degree to which observed change is actual variation in performance rather than error in the measurement process (10).

The most commonly employed research tool seeking to measure movement competency within athletic populations is the Functional Movement Screen (FMS) – a battery of seven movements each rated on an ordinal scale from 0-3 (18). Current evidence suggests that FMS

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

component movement scores, along with the combined total score both exhibit acceptable inter-rater reliability when scored in real-time and from video. This is indicated by weighted kappa coefficients ($k_w$) of 0.31-1.00 for component movements and intraclass correlation coefficients (ICC) of 0.89-0.99 for the total FMS score (13, 18, 21, 22, 26). Further, it appears that when raters undertake the FMS certification, the screening process can be applied reliably regardless of experience (15). The intra-rater reliability varies to a greater extent, with ICC values for the total FMS score ranging between 0.60 and 0.96 and $k_w$ coefficients of between 0.16 and 1 (7, 21, 22, 25, 26).

While current evidence suggests the FMS is a tool which possesses acceptable inter- and intra-rater reliability, its use in assessing AMC may be limited. The movement patterns which comprise the FMS are intentionally basic, leading to a tool lacking the complexity to reflect the structural demands of dynamic athletic movement (23). Furthermore, the FMS uses a basic four-point scale to address pain and asymmetry, and seeks only to dichotomise movement based on the presence or otherwise of global dysfunction (2). Consequently, research has shown that individuals who receive the same FMS score on a particular movement pattern may perform the movement in very different ways (5). Thus, while the FMS provides practitioners with a basic screen to highlight gross movement dysfunction, there remains scope for a more advanced and specific level of movement assessment. Any additional level of movement analysis must address the specific competencies required to undertake the more complex and dynamic movements which are associated with athletic training.

Several authors have proposed movement assessments which may assist in the understanding and development of athletic movement to a greater extent than the FMS (11, 17, 19, 23, 24, 28). By assessing movements specific to the training demands of the athlete(s)

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

concerned, this new research acknowledges the need for an additional level of assessment, above that of a general screen. Furthermore, some of these tools use explicit scoring criteria to highlight the specific region limiting the proficiency of a movement (17, 19). One such assessment, the Athletic Ability Assessment (AAA) was originally presented by McKeown et al. (17) and consisted of 7 movements, each possessing 3 key assessment areas which were rated on an ordinal scale from 1-3 (3 perfect; 2 OK; 1 poor). The AAA has since been adapted for use in field-based team sports such as Australian Rules football (ARF) (32). The protocol presented in this adaptation narrowed the assessment down to 5 movements but maintained the scoring method used by McKeown et al. (17). Using this 5-movement version of the AAA, all component movements exhibited moderate to substantial inter-rater reliability when scored retrospectively from two-dimensional video ($k_w$=0.59-0.77) (32). While suggestive of acceptable inter-rater agreement, this study only employed two raters, both with similarly high levels of professional experience. Thus, the effect of rater experience on reliability of assessment remains unexplored. Additionally, no intra-rater analysis was performed and scoring was undertaken only via video. If a movement assessment is to be of use in tracking changes in AMC at all levels of sport, its test-retest scoring stability, when assessed in real-time, and by inexperienced raters must be known. Therefore, the purpose of this study was to assess the inter- and intra-rater reliability of the AAA when assessed from video and in real-time, in a cohort of sub-elite ARF players. In addition, the study sought to determine the effect of professional experience on rater reliability.

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

## METHODS

### *Experimental Approach to the Problem*

Subjects undertook the AAA on two occasions separated by one week. Both testing sessions were performed between 3 and 6 pm as a warm-up prior to the subjects' weekly gym session. Subjects were excluded from the second assessment if they incurred any injury which warranted missed training time during the intervening week or reported pain at any time during either assessment. All subjects had performed the AAA component movements as part of their training program prior to taking part in the study and were provided standardised instructions for the completion of each movement. Due to the researchers' inability to control preceding football sessions or subjects' work and/or study commitments, no attempt was made to standardise exercise during the days prior to testing.

During the first testing session, all participants were filmed in the frontal and sagittal plane. Ten raters (one experienced and nine novices) were assigned in a quasi-random manner to complete either (a) real-time assessment on two occasions to determine real-time intra-rater reliability, (b) real-time assessment on one occasion to determine real-time inter-rater reliability or (c) video-based assessment on two occasions to determine video-based inter- and intra-rater reliability.

### *Subjects*

Eighteen sub-elite male ARF players were convenience sampled from one South Australian National Football League (SANFL) club. Of this sample, 13 players were part of the club's Academy squad which played in the SANFL reserves competition (age 20.2 ± 1.0 yrs; body mass 85.6 ± 28.6 kg; height 173.6 ± 28.4 cm) and five were contracted players eligible to

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

play in the SANFL League competition (age 22.9 ± 3.2 yrs; body mass 81.3 ± 6.7 kg; height 184.5 ± 6.6 cm). All participants in the study were deemed free of any musculoskeletal injury by club physiotherapists and were undertaking full training duties at the time. The study was approved by the University of South Australia Human Research Ethics Committee and all subjects were informed of the benefits and risks associated with participation in the study before providing written informed consent.

*Raters*: Ten raters took part in the study, nine of whom (3 females and 6 males) were university students enrolled in the programs of Human Movement, Physiotherapy and Clinical Exercise Physiology at the University of South Australia, and acted as novice raters ($NOV_{1-9}$). The novice raters had never used the AAA prior to their participation in the study. The lead author on this study acted as the experienced rater (EXP), having used the AAA with elite and sub-elite ARF players for the previous 2 years, and worked to develop the scoring criteria used. All novice raters had no previous contact with participants in the study meaning they were blind to their clinical information, previous AAA scores and any additional cues which may affect scoring decisions. Conversely, EXP had worked as a strength and conditioning coach with the participants for 2 years and was not blind to these factors.

**Procedures**

*The Athletic Ability Assessment (AAA)*: The AAA consists of 5 movements which underlie the foundational gym-based exercises used to improve muscular strength, power and coordination. These movements include the overhead squat, double lunge, single leg Romanian deadlift, chin-up and push-up (Figure 1). As was the case in the study by Woods et al. (32), each movement possessed three key scoring areas, each relating to a specific aspect of the movements performance. Each key scoring area was rated on an ordinal scale from 1-3 (3 perfect; 2 OK; 1

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

poor), resulting in a total score of between 3 and 9 for each movement. When movements were performed bilaterally, a separate score was assigned for each side, resulting in a total score for the movement assessment of between 21 and 63. However, researchers in this study sought to clarify the specific wording of the criteria used by Woods et al (32), and in some cases, change the focus of the key scoring areas for each movement. Furthermore, to reflect the interrelatedness and complexity of human movement the score provided in some key areas was made dependent on the that obtained in others. The dynamic systems theory of human movement suggests that a movement pattern will vary greatly depending on the physical constraints of the individual and the constraints applied to the movement task (30). In the context of a movement assessment, raters are 'scoring' a movement pattern which has self-organised based upon the physical constraints specific to the subject (i.e. strength, balance, mobility) and the task constraints applied by the rater (movement instructions). For example, in the presence of poor hip and/or ankle mobility (physical constraints), a subject may reduce squat depth markedly (a major compensation) to avoid forward inclination of the torso and arms, thus maintaining the dowel above head and feet (task constraint). To capture this phenomenon, changes were made to the criteria such that if a major compensation was present in one key area (a rating of 1), the rater was directed to change any key areas rated as 3 (good), to 2 (meets the criteria for a 3 but does so with a major compensation elsewhere in the movement). The scoring criteria used in the study can be seen in Table 1.

All movements were performed for 5 repetitions, except the chin-up and push-up which had repetition targets of 10 and 30 respectively, to assess the effect of fatigue on the key areas assessed in those movements. All participants performed the assessment unshod and wore sports shorts only to allow viewing of the torso and scapulae during assessment. The movements were

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

performed in the following order: overhead squat, double lunge left (left [L]), double lunge right (right [R]), single leg Romanian deadlift left (L), single leg Romanian deadlift right (R), chin-up and push-up.

*****FIGURE 1 ABOUT HERE*****

*****TABLE 1 ABOUT HERE*****

*Rater Training*: Before undertaking the assessment process, the novice raters underwent a 2-hour education session, delivered by the lead author (EXP). This session presented the rationale behind the AAA along with the interpretation and application of its scoring criteria. Raters were then given the chance to practice using the assessment and performing its component movements.

*Assessment of Real-Time Reliability:* The experienced rater (EXP), along with two novice raters ($NOV_1$ and $NOV_2$) performed all assessments during both real-time testing sessions. Alongside these raters, an additional 3 novice raters ($NOV_3$, $NOV_4$ and $NOV_5$) assessed participants during the second real-time testing session. Individual intra-rater reliability for EXP, $NOV_1$ and $NOV_2$ was determined by comparing their scoring decisions during the first and second real-time testing sessions. The scores provided by all 6 raters (EXP and $NOV_{1-5}$) during the second real-time testing session were compared to determine real-time inter-rater reliability.

*Assessment of Video-Based Reliability:* While being observed by the raters during real-time session one, each subject was individually filmed from the frontal and sagittal plane using two Canon Legria FS200 video cameras. The camera filming in the frontal plane for the chin-up and push-up was positioned behind the participants so that viewing of the scapulae was possible during these movements. The footage was then edited to produce individual files for all 18

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

subjects, featuring each movement performed simultaneously in both planes, on a split screen. These video files were then provided to EXP and the remaining 4 novice raters ($NOV_{6-9}$), who completed video-based assessment on two occasions. Raters were allowed two weeks to assess all video files with a one-week period between repeat assessments. Raters were permitted to view footage multiple times and footage could be slowed if deemed necessary. The individual intra-rater reliability for each rater was determined by comparing their scoring decisions during video-based session 1 to those provided in video-based session 2. Video-based inter-rater reliability was determined by comparing the scoring decisions from each rater during session 1.

***Statistical Analyses***

Statistical analyses took place using SPSS (IBM SPSS Inc, Chicago Version 23) and Analyse-it statistical software (Method Validation Edition version 4.0). The reliability of all AAA total scores was determined using a 2-way random intraclass correlation coefficient ($ICC_{2,1}$). An $ICC_{2,1}$ was also applied to determine measures of inter-rater reliability due to its capability to deal with fully crossed study designs using more than one rater, as was the case in this study (8). $ICC_{2,1}$ scores were interpreted as follows: high reliability, 0.90–0.99; good reliability, 0.80–0.89; fair reliability, 0.70–0.79; poor reliability, 0.00–0.69 (21).

The intra-rater agreement of all key areas and component movements was assessed using a weighted kappa coefficient ($k_w$). Kappa is a chance-corrected measure of agreement, meaning it accounts for the level of agreement which would be expected if all ratings were assigned in a simple random manner (1, 31). The weighted kappa was chosen to account for the magnitude of disagreement displayed by raters, with a quadratic weighting scheme applied in this instance (20). The $k_w$ coefficient was interpreted as follows: perfect, 1.0; almost perfect, 0.81-0.99; substantial,

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

0.61-0.80; moderate, 0.41-60; fair, 0.21-0.40; slight, 0.01-0.20; less than chance agreement, <0.01 (12).

When assessing agreement using the $k_w$ coefficient, any tendency for raters to distribute scores more frequently into a single category can result in low or incalculable coefficients despite high absolute agreement (1, 4). For this reason, absolute percentage agreement was also calculated for key assessment areas and component movements to contextualise $k_w$ results.

Intra-rater typical error (TE) and minimum detectable change (MDC) was calculated for each component movement and total score. The following equation was used to calculate the TE (9):

$$St\ dev.(change\ scores)/\sqrt{2}$$

The MDC was then calculated as follows (9):

$$1.96\ x\ TE\ x\ \sqrt{2}$$

All data was presented as mean ± 95% confidence intervals (CI), unless otherwise stated.

## RESULTS

*Intra-Rater Reliability:* Of the 21 key assessment areas rated in real-time, moderate or greater agreement was exhibited on 11 occasions by EXP, 7 occasions by $NOV_1$ and 3 occasions by $NOV_2$ (Table 2). Key assessment areas for the single leg Romanian deadlift on both sides showed poorest overall agreement. Real-time intra-rater agreement for component movements ranged from slight to almost perfect (Table 3). Of the 7 component movements, moderate or greater intra-rater agreement was exhibited on 5 occasions by both EXP and $NOV_1$, and on 2 occasions by $NOV_2$. The single leg Romanian deadlift on both sides showed poorest intra-rater

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

agreement, with only EXP exhibiting greater than fair agreement for either the left or right side of the movement. The $ICC_{2,1}$ for total score ranged between 0.50 and 0.61, indicating poor intra-rater reliability for this variable when assessed in real-time.

*****TABLE 2 ABOUT HERE*****

*****TABLE 3 ABOUT HERE*****

When assessed by video, of the 21 key assessment areas, moderate or greater agreement was exhibited on 18 occasions by EXP and $NOV_7$, 13 occasions by $NOV_8$ and $NOV_9$, and 12 occasions by $NOV_6$ (Table 4). For component movements, video-based intra-rater agreement was predominantly moderate to almost perfect for all raters (Table 5). Specifically, of the 7 component movements, moderate or greater agreement was exhibited on 7 occasions by $NOV_6$ and $NOV_9$ and 6 occasions by EXP, $NOV_7$ and $NOV_8$. Only EXP showed high intra-rater reliability for total score when assessed by video ($ICC_{2,1}$=0.93), with all other raters exhibiting poor to moderate reliability for this variable ($ICC_{2,1}$=0.60-0.73).

*****TABLE 4 ABOUT HERE*****

*****TABLE 5 ABOUT HERE*****

The intra-rater TE and MDC are shown for each component movement and total score in Tables 6 and 7. $NOV_1$ exhibited the best real-time score stability, producing the smallest MDC for total score and four of seven component movements. EXP and $NOV_6$ exhibited the most consistent score stability during video-based scoring. Specifically, EXP produced the smallest MDC for total score and two of seven component movements while $NOV_6$ produced the smallest MDC for three of seven component movements.

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

*****TABLE 6 ABOUT HERE*****

*****TABLE 7 ABOUT HERE*****

*Inter-Rater Reliability:* Inter-rater reliability of key assessment areas was poor when assessed via both modalities (Table 8). Of the 21 key assessment areas, there were only two for real-time assessment and one for video-based assessment that were not classified as possessing poor reliability. All component movements, except the chin-up when assessed in real-time, showed poor inter-rater reliability (Table 9). For both assessment modalities, the inter-rater reliability of AAA total score was classified as poor (real-time $ICC_{2,1}$=0.52; video-based $ICC_{2,1}$=0.47).

*****TABLE 8 ABOUT HERE*****

*****TABLE 9 ABOUT HERE*****

## DISCUSSION

The findings of this study suggest that the AAA possesses acceptable intra-rater reliability when assessment is undertaken via video and especially so if the rater is experienced using the tool. Given the variable intra-rater agreement levels, caution should be taken if assessment is performed in real-time. Furthermore, appropriate measures should be taken before AAA results are compared between raters due to the overall poor inter-rater agreement exhibited by the raters in this study.

*Real-Time Intra-Rater Reliability:* Rater agreement on key assessment areas of the AAA was inconsistent. Double lunge hip range on both sides, along with the push-up and chin-up rep targets exhibited high levels of agreement, while all key assessment areas of the single leg Romanian deadlift (L and R) showed predominantly slight to fair agreement.

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

Assessment of push-up and chin-up rep targets was an objective task, explaining why both criteria exhibited mostly substantial to perfect agreement. Conversely, to assess the single leg Romanian deadlift, raters were required to determine hip rotation along the sagittal axis, hinge range, as well as single leg control. Thus, the predominantly slight to fair agreement on all key assessment areas of this movement may be due to the requirement for raters to view it from multiple angles in a relatively short period (5 repetitions). Researchers interested in the reliability of the FMS have raised similar issues regarding the hurdle step and rotary stability test (21, 22). Both tests require assessing movement in multiple planes and are often the two movements which exhibit the poorest intra-rater agreement among the FMS testing battery (21, 22, 29). Such an issue may have been exacerbated in the current study by the relatively high number of raters used for real-time assessment. This could have limited the range of angles from which each rater could freely view subjects, particularly during session 2.

When assessed by EXP and $NOV_1$, AAA component movements exhibited comparable real-time intra-rater agreement to other commonly used movement screens such as the FMS (21, 29). Interestingly, $NOV_1$ exhibited the greatest intra-rater agreement, registering substantial to almost perfect agreement on 5 of 7 component movements (compared to 3 of 7 and 2 of 7 for EXP and $NOV_2$ respectively). This greater level of agreement for a novice rater when compared to EXP initially appears counterintuitive. However, such an occurrence may be explained with reference to previous research, along with a brief analysis of the statistical theory surrounding the *k* coefficient. Firstly, as noted by Smith et al. (26), the assessment of real-time intra-rater reliability assumes that subjects performed the test in an identical manner on two occasions. While researchers took precautions to ensure this was the case, a change in movement quality from real-time testing session 1 to 2 cannot be discounted. If EXP was better able to identify

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

such changes - as would be expected considering their familiarity with the subjects - this would have reduced their agreement levels, despite greater precision of assessment. Secondly, the *k* coefficient is corrected for agreement expected by chance, and is represented as follows (1, 3, 18):

$$Po - Pe / 1 - Pe$$

Here, $P_o$ = absolute agreement and $P_e$ = expected agreement. Evidently, if $P_e$ is high, the *k* coefficient will be low for any given level of $P_o$ - a problem known as 'prevalence' (1, 4). Prevalence occurs when a rater distributes a large proportion of their ratings into a single category (1, 4). In the current study, it appears that this occurred on several occasions as seen in Table 2 where the $k_w$ was either not calculable, despite 100% absolute agreement, or *0*, despite relatively high absolute agreement. Interestingly this issue of prevalence was more common in the novice raters and tended to occur when assessing key areas that required careful positioning close to the subject (i.e. chin-up scapular rhythm and push-up upper quadrant). Thus, while novice raters tended to exhibit greater real-time agreement on total score and most component movements, this may have been due to chance agreement resulting from a tendency to guess during the assessment of harder-to-view key assessment areas. As stated, this study employed a higher number of raters during real-time assessment than other similar research studies, and in doing so may have encouraged novice raters to guess by limiting their free movement around the subjects. However, as EXP did not experience this problem of prevalence in their scoring decisions, other factors such as decreased motivation may have played a part. As both real-time testing sessions lasted 3 hours, a lack of experience assessing movement over this period may have limited the desire of novice raters to move around the subjects and assess as accurately as possible.

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

The MDC scores presented in this study have important practical implications for those applying the AAA as they may assist practitioners in identifying meaningful change. $NOV_1$ registered the smallest MDC for 4 of 7 component movements and $NOV_2$, 3 of 7 component movements. These results, however, must be interpreted in the context of $k_w$ results. For example, $NOV_2$ possessed the smallest real-time MDC for the double lunge (L) movement (0.91), yet their $k_w$ coefficient for this movement was 0.09, representing only slight agreement. Further, 2 of 3 double lunge (L) key assessment areas exhibited less than chance agreement when assessed by $NOV_2$ - a result which appears to be due to prevalence in their scoring decisions. In this instance, a low MDC seems irrelevant if the component movement in question possesses slight or less than chance agreement.

In practice, coaches using the AAA may need to apply an 'agreement cut-off' to legitimise MDC scores. For example, a component movement may need to register at least moderate agreement ($k_w$>0.40) before MDC scores can be considered. Ideally, a component movement meets both a pre-determined agreement cut-off along with an MDC of less than 1 point, so that any individual change in score can be considered real. The double lunge (L), double lunge (R) and push-up all met this criterion, while the chin-up, single leg Romanian deadlift (R) and overhead squat, with MDC scores of 1.39, 1.83 and 2.61, did not. The single leg Romanian deadlift (R) did not possess the agreement cut-off of $k_w$>0.40, when assessed by any rater for its MDC to be considered. Employing a pre-determined agreement cut-off along with an understanding that interpretation of change may require a more conservative approach for some component movements will ensure coaches apply the AAA in the most accurate way possible.

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

The real-time MDC values for total score ranged between 5.29 ($NOV_1$) and 8.07 (EXP). As the AAA total score represents a 42-point scale, this equates to 13%, 14% and 19% of the total score, for $NOV_1$, $NOV_2$ and EXP respectively. For the novice raters, these real-time MDC scores are similar to the 10% presented by Teyhen et al. (29) for the FMS total score. However, AAA total score possessed poor reliability when assessed by all raters in real-time ($ICC_{2,1}$<0.70). As discussed, MDC values should be applied cautiously if the variable in question does not meet a predetermined reliability or agreement cut-off. Should coaches decide to apply the AAA in a field setting and use the total score to track changes in AMC in their athletes, they should first endeavor to obtain fair intra-rater reliability (ICC>0.69) for this variable before using intra-rater MDC scores to interpret change.

*Video-based intra-rater reliability:* When assessed by video, an improvement in the intra-rater agreement of key assessment areas was apparent. Double lunge hip range on both sides again showed consistently substantial to perfect agreement for all raters. Notably, EXP showed improved intra-rater agreement on key assessment areas of the single leg Romanian deadlift when assessing from video. This lends support to the suggestion that poor real-time agreement levels for this movement are in part related to an inherent difficulty in assessing it under real-time conditions. This said, the $k_w$ for the single leg Romanian deadlift remained the lowest of all movements during video-based assessment, implying that other factors such as ambiguous scoring criteria may have added to the apparent difficulty in assessing this movement.

Ambiguity in scoring criteria is one reason suggested for the decreased agreement of FMS component movements such as the hurdle step and in-line lunge (18, 22). Parenteau et al. (22) stated that for these movements, a score of 2 out of 3 is given if the subject exhibits an *acceptable* compensation, while *too-much* compensation results in a score of 1 out of 3. These

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

terms may increase the subjective nature of scoring decisions which possibly leads to increased variability. In the same way, the use of *slight* and *excessive* in the AAA scoring criteria may have led to the poorer intra-rater agreement of component movements such as the single leg Romanian deadlift and overhead squat. For example, the scoring criteria for overhead squat knee alignment directs the rater to assign a 2 for *any slight deviation* and a 1 for *any excessive deviation*. Further, if a rater were to assign a 1 for knee alignment and a 3 for the other two key areas of the overhead squat, they would be directed to change both scores of 3 to 2, due to the presence of a major dysfunction (i.e. a 1 for knee alignment). In this instance, if a rater interpreted the terms *slight* and *excessive* differently during repeat testing sessions it would be the difference between a score of 8 and a score of 5 for the overhead squat. Thus, it is apparent that a relatively small difference in interpretation can lead to a large difference in the overall score for a component movement. This, highlights the importance of rater education prior to use of movement assessments such as the AAA so that an understanding of what constitutes a major dysfunction is clear and can be interpreted consistently.

While, for some raters, the single leg Romanian deadlift and overhead squat were difficult to assess by video it should be noted that on only 3 occasions did intra-rater agreement drop below moderate for any component movement, across all raters. Considering five raters assessed 7 movements each, resulting in 35 separate measures of agreement, this finding augurs well for the use of the AAA when assessed by video. For EXP, especially, the video-based $k_w$ coefficients are similar to those of the FMS component movements presented by Parenteau et al. (22), which ranged between 0.65 and 0.98. Aside from the single leg Romanian deadlift (left), the intra-rater agreement levels for EXP are also in line with those presented by Woods et al. (33) who found substantial to almost perfect intra-rater agreement for an experienced rater using

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

the AAA. Furthermore, when assessed by EXP, the AAA total score exhibited an MDC of 2.99 or 7%. Importantly, this reduced MDC was recorded with an accompanying increase in the reliability of the total score which reached an $ICC_{2,1}$ of 0.93 when assessed by video. This overall increase in the intra-rater agreement of component movement scores along with a dual increase in the absolute and relative reliability of total score when assessed by EXP, suggests the intra-rater score stability of the AAA is maximised when assessed by an experienced rater from video.

As is evident from Tables 6 and 7, the MDC ranges markedly between raters for most component movements and total score. When assessed by video, EXP can be confident that a change of 3 points in the AAA total score can be considered worthwhile. However, if $NOV_9$ performs the assessment, a change of 9 points would be required to make definitive conclusions regarding changes in performance. These findings highlight the importance of determining the individual reliability for any rater applying the AAA as it is evident that the raters in this study exhibited distinct psychometric properties.

*Inter-Rater Reliability*: Other than the key assessment areas of push-up and chin-up rep targets, along with the chin-up component movement, all other AAA variables exhibited poor inter-rater reliability when assessed via both modalities. The reason for this apparent poor inter-rater reliability is likely due to a combination of factors including varied viewing angles during real-time session 2, ambiguous scoring criteria, within-subject movement variability and/or insufficient rater education.

As previously stated, with 6 raters present during real-time session 2, the range of angles from which each movement could be assessed was probably limited. However, video-based inter-rater reliability was also poor despite allowing all raters a standardised view of each subject. This suggests that differences in raters' interpretation of the scoring criteria may have been a

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

greater contributor to the poor inter-rater reliability exhibited in this study. As discussed, interpreting terms such as *slight* and *excessive* may have played a role in reducing intra-rater reliability. In fact, it is more likely that the interpretation of such terms played a role in the poor inter-rater reliability, as it is reasonable to assume that differences in interpretation were greater between than within raters. Furthermore, some criterion required a dichotomous scoring decision (i.e. 3 = stick remained above head and feet / 1 = stick did not remain above head and feet). In sub-elite athletes, it is reasonable to assume that that some movement inconsistencies were present. If this was the case, some subjects may have exhibited movement characteristics which met the criteria for both a '1' and a '3', within the space of five repetitions. If some raters were more influenced by the presence of a single very poor (or very good) repetition, this may have contributed to a certain level of inter-rater disagreement. When using the FMS, raters are instructed to use the lowest score of three repetitions (7). Similarly, to obtain a score of '3' for AAA key areas, it is intended that subjects meet the corresponding criteria for all repetitions. While this is indicated in some criteria (i.e. does the subject exhibit *any* compensation), some lack explicit instruction in this area. Future research on the AAA should seek to clarify this by way of altered scoring criteria and/or more thorough rater education.

It was an aim of this study to employ several novice raters so that findings could be applied to those who work at lower levels of sport. All novice raters were students of either Human Movement, Physiotherapy or Clinical Exercise Physiology and had never used the AAA prior to participating in the study. As such, a 2-hour education session was used to familiarise the novice raters with the philosophy and principles behind the AAA and provide them an opportunity to practice the assessment in an informal setting. While all novice raters had a similar level of formal education, their level of experience assessing movement in professional

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

settings, and consequently their knowledge of the underlying principles of movement assessment was unknown. Therefore, it may be that the education process did not properly reinforce the principles behind the assessment of the AAA component movements, and the novice raters' level of professional experience was insufficient to counter this.

## PRACTICAL APPLICATION

The results of this study show that the AAA is most reliably assessed retrospectively by video, with this assessment modality showing predominantly substantial to almost perfect intra-rater agreement. This is an important finding for coaches as it means multiple subjects can be filmed simultaneously, expediting the data collection process. Considering the widespread availability of electronic devices such as smart phones and iPads this appears a viable option if sufficient space is available.

While acceptable levels of video-based intra-rater agreement were evident across a range of raters, the inter-rater reliability of the AAA was poor. It is recommended that the practical application of the AAA within sporting organisations be preceded by a stringent training process so that the interpretation of its scoring criteria is standardised across raters within whom comparison is necessary.

Lastly, this study shows that intra-rater agreement can vary widely between raters for the same component movement. Therefore, along with a stringent training process prior to its application, practitioners and researchers should determine the individual psychometric properties for all raters prior to applying the AAA.

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

## ACKNOWLEDGMENTS

This research was part-funded by the Port Adelaide Football Club. Researchers wish to acknowledge players from the Port Adelaide Football Club (SANFL) and students from the University of South Australia for their participation in the research project.

## REFERENCES

1. Byrt, T, Bishop, J and Carlin, JB. Bias, prevalence and kappa. *J Clin Epidemiol* 46: 423-429, 1993.
2. Cook, G, Burton, L and Hoogenboom, B. Pre-participation screening: the use of fundamental movements as an assessment of function - part 1. *N Am Sports Phys Ther* 1: 62-72, 2006.
3. Di Eugenio, B and Glass, M. The kappa statistic: a second look. *Comput Linguist* 30: 95-101, 2004.
4. Feinstein, AR and Cicchetti, DV. High agreement but low kappa: I. the problems of two paradoxes. *J Clin Epidemiol* 43: 543-549, 1990.
5. Frost, DM, Beach, TA, Campbell, TL, Callaghan, JP and McGill SM. An appraisal of the functional movement screen grading criteria – is the composite score sensitive to risky movement behavior? *Phys Ther Sport* 16: 324-330, 2015.
6. Gamble, P. Movement screening protocols: rationale versus evidence. *New Zealand J Sports Med* 40: 83-87, 2013.
7. Gribble, PA, Briggle J, Pietrosimone, BG, Pfile KR and Webster, KA. Intrarater reliability of the functional movement screen. *J Strength Cond Res* 27: 978-981, 2013.
8. Hallgren, KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol* 8: 23-34, 2012.
9. Hopkins, W, Spreadsheets for analysis of validity and reliability. *Sportscience* 2015, available at: www.sportsci.org/2015/ValidRely.htm
10. Hopkins, WG. How to interpret changes in an athletic performance test. *Sportscience* 8: 1-7, 2004.

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

11. Kritz, M. Development, reliability and effectiveness of the movement competency screen (MCS). Auckland University of Technology: Auckland, NZ, 2012.
12. Landis, JR and Koch, GG. The measurement of observer agreement for categorical data. *Biometrics* 33: 159-174, 1977.
13. Leeder, JE, Horsley, IG and Herrington, LC. The inter-rater reliability of the functional movement screen within an athletic population using untrained raters. *J Strength Cond Res* 30: 2015: 2591-2599.
14. Lloyd, RS and Oliver JL. The youth physical development model: a new approach to long-term athletic development. *Strength Cond J* 34: 61-72, 2012.
15. McCunn, R, aus der Funten,K, Fullager, HK, McKeown, I and Meyer, T. Reliability and association with injury of movement screens: a critical review. *Sports Med* 1-19, 2015.
16. McKeown, I and Ball, N. Current practices of long term athlete development of junior athletes in high performance sports environments. *J Aust Strength and Cond* 21: 4-12, 2013.
17. McKeown, I, Taylor-McKeown, K, Woods, C and Ball, N. Athletic ability assessment: a movement assessment protocol for athletes. *Int J Sports Phys Ther* 9: 862-873, 2014.
18. Minick, KI, Kiesel, KB, Burton, L, Taylor, A, Plisky, P and Butler, RJ. Interrater reliability of the functional movement screen. *J Strength Cond Res* 24: 479, 2010.
19. Mischiati, CR, Comerford, M, Gosford, E, Swart, J, Ewings, S, Botha, N, Stokes M and Mottram SL. Intra and inter-rater reliability of screening for movement impairments: movement control tests from the foundation matrix. *J Sport Sci Med* 14: 427-440, 2015.
20. Norman, GR and DL Streiner, *Biostatistics: the bare essentials*. Hamilton, BC Decker, 2008.
21. Onate, JA, Dewey, T, Kollock, RO, Thomas, KS, Van Lunen, BL, Demaio, M and Ringleb, SI. Real-time intersession and interrater reliability of the functional movement screen. *J Strength Cond Res* 26: 408-416, 2012.


ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

22. Parenteau, E, Gaudreault, N, Chambers, S, Boisvert, C, Grenier, A, Gagne, G and Balg, F. Functional movement screen test: a reliable screening test for young elite ice hockey players. *Phys Ther Sport* 15: 169-175, 2013.
23. Parsonage, JR, Williams, RS, Rainer, P, McKeown, I and Williams, MD. Assessment of conditioning specific movement tasks and physical fitness measures in talent identified under 16-year-old rugby union players. *J Strength Cond Res* 28: 1497-1506, 2014.
24. Reid, DA, Vanweerd, RJ, Larmer, PJ and Kingstone R. The inter and intra rater reliability of the netball movement screening tool. *J Sci Med Sport* 18: 353-357, 2015.
25. Shultz, R, Anderson, SC, Matheson, GO, Marcello, B and Besier, T. Test-retest and interrater reliability of the functional movement screen. *J Athl Train* 48: 331-336, 2013
26. Smith, CA, Chimera, NJ, Wright, NJ and Warren MJ. Interrater and intrarater reliability of the functional movement screen. *J Strength Cond Res* 27: 982-987, 2013.
27. Suchomel, TJ, Nimphius, S and Stone, MH. The Importance of muscular strength in athletic performance. *Sports Med* 46: 1419–1449, 2016.
28. Tarara, DT, Hegedus, EJ and Taylor, JB. Real-time test-retest and interrater reliability of select physical performance measures in physically active college-aged students. *Int J Sports Phys Ther* 9: 874-887, 2014.
29. Teyhen, DS, Shaffer, SW, Lorenson, CL, Halfpap, JP, Donofry, DF, Walker, MJ, Dugan JL and Childs, JD. The functional movement screen: a reliability study. *J Orthop Sports Phys Ther* 42: 530-540, 2012.
30. Thelen, E. Dynamic systems theory and the complexity of change. *Psychoanal Dialogues* 15: 255-283, 2005.
31. Viera, AJ and Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 37: 360-363, 2005
32. Woods, CT, McKeown, I, Haff, GG and Robertson, S. Comparison of athletic movement between elite junior and senior Australian football players. *J Sports Sci* 24: 1260-1265, 2015.

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

548 33. Woods, CT, Banyard, HG, McKeown, I, Fransen, J and Robertson, S. Discriminating talent
549 identified junior Australian footballers using a fundamental gross athletic movement assessment.
550 *J Sport Sci Med* 15: 548-553, 2016.

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

**Table 1**: Qualitative scoring criteria corresponding to each performance level, for each AAA component movement

| | KEY AREA | GOOD - 3 | OK - 2 | BAD - 1 |
|---|---|---|---|---|
| **Overhead Squat** | **Knee Alignment** | In line with toes | Any slight deviation OR '3' with a '1' elsewhere | Any excessive deviation |
| | **Upper Quadrant** | Hands above head and feet | '3' with a '1' elsewhere | Unable to achieve position |
| | **Triple Flexion** | Parallel with neutral lumbar | Parallel with loss of neutral OR '3' with a '1' elsewhere | Unable to reach parallel |
| **Double Lunge** | **Hip, Knee, Ankle** | Consistent alignment on stance leg | Any slight deviation | Any excessive deviation |
| | **Take Off Control** | Controlled constant motion | Any slight loss of balance or jerking movement OR '3' with a '1' in hip range | Any excessive loss of balance |
| | **Hip Range** | Able to obtain knee to ground | '3' with '1' in take-off control | Unable to obtain knee to ground |
| **Single Leg Romanian Deadlift** | **Single Leg Control** | Controlled constant motion throughout | Slight loss of balance or jerking movement | Excessive loss of balance |
| | **Hip Control (Sagittal)** | No rotation | Any slight rotation OR '3' with a '1' elsewhere | Any excessive rotation |
| | **Hinge Range** | Neutral lumbar and parallel | Neutral but not parallel OR '3' with a '1' in single leg control | Unable to dissociate hips and trunk |
| **Push-Up** | **Total Body Control** | Perfect alignment all reps | Loss of alignment some reps | Unable to align at all |
| | **Upper Quadrant** | No scapula winging or elevation | Loss of position for some reps | Unable to obtain position |
| | **X30 reps** | Yes | '3' with a '1' elsewhere | No |
| **Chin-Up** | **Scapular Rhythm** | No scapula winging or elevation | Loss of position for some reps | Unable to obtain position |
| | **Total Body Control** | Perfect alignment | Loss of alignment for some reps | Unable to obtain alignment |
| | **X10 reps** | Yes | '3' with a '1' elsewhere | No |

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

**Table 2**: Real-time intra-rater agreement of key assessment areas

| | | EXP | | $NOV_1$ | |
|---|---|---|---|---|---|
| | | **$k_w$ (95% CI)** | **%Agree.** | **$k_w$ (95% CI)** | **%Agre** |
| | Knee Alignment | 0.14 (-0.24 - 0.53) | 55.6 | 0.67 (0.32 - 1.00) | 83.3 |
| **OHS** | Upper Quadrant | 0.22 (-0.21 - 0.65) | 55.6 | 0.36 (-0.07 - 0.78) | 66.7 |
| | Triple Flexion | 0.21 (-0.23 - 0.65) | 61.1 | 0.62 (0.29 - 0.96) | 66.7 |
| | Hip, Knee, Ankle | 0.18 (-0.22 - 0.58) | 50.0 | na | 100.0 |
| **DLL** | Take Off Control | 0.72 (0.47 - 0.97) | 83.3 | 0.31 (-0.28 - 0.90) | 83.3 |
| | Hip Range | 0.00 (0.00 - 0.00) | 94.4 | 0.79 (0.77 - 0.81) | 94.4 |
| | Hip, Knee, Ankle | 0.70 (0.41 - 0.98) | 77.8 | 0.00 (0.00 - 0.00) | 94.4 |
| **DLR** | Take Off Control | 0.30 (-0.10 - 0.69) | 44.4 | 0.77 (0.34 - 1.00) | 94.4 |
| | Hip Range | 0.44 (-0.21 - 1.00) | 88.9 | 1.00 (1.00 - 1.00) | 100.0 |
| | Single Leg Control | 0.18 (-0.25 - 0.61) | 66.7 | -0.19 (-0.39 - 0.02) | 61.1 |
| **SLRDLL** | Hip Control (S) | 0.07 (-0.01 - 0.15) | 83.3 | 0.29 (-0.11 - 0.69) | 66.7 |
| | Hinge Range | 0.47 (0.04 - 0.89) | 77.8 | -0.08 (-0.19 - 0.03) | 83.3 |
| | Single Leg Control | 0.45 (0.10 - 0.80) | 50.0 | 0.14 (-0.27 - 0.55) | 55.6 |
| **SLRDLR** | Hip Control (S) | 0.29 (-0.17 - 0.75) | 77.8 | 0.36 (-0.03 - 0.74) | 66.7 |
| | Hinge Range | 0.69 (0.29 - 1.00) | 83.3 | -0.08 (-0.19 - 0.03) | 83.3 |
| | Total Body Control | 0.40 (-0.04 - 0.84) | 72.2 | 0.00 (0.00 - 0.00) | 77.8 |
| **PU** | Upper Quadrant | 0.51 (0.30 - 0.72) | 55.6 | 0.00 (0.00 - 0.00) | 94.4 |
| | X30 Reps | 0.48 (-0.09 - 1.00) | 77.8 | 0.64 (0.00 -1.00) | 94.4 |
| **CU** | Scapular Rhythm | 0.45 (0.11 - 0.79) | 66.7 | 0.00 (0.00 - 0.00) | 94.4 |

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

| | | | | |
|---|---|---|---|---|
| Total Body Control | 1.00 (1.00 - 1.00) | 100.0 | 0.22 (-0.30 - 0.74) | 77.8 |
| X10 Reps | 0.78 (0.49 - 1.00) | 88.9 | 0.66 (0.31 - 1.00) | 83.3 |

OHS = overhead squat; DLL = double lunge left; DLR = double lunge right; SLRDLL = single leg Romanian deadlift
deadlift right; CU = chin-up; PU = push-up; S = sagittal axis

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

**Table 3**: Real-time intra-rater agreement of component movements and total score

| | EXP | | $NOV_1$ | | $NOV_2$ | |
|---|---|---|---|---|---|---|
| | $k_w$ (95% CI) | %Agree. | $k_w$ (95% CI) | %Agree. | $k_w$ (95% CI) | %Agree. |
| **OHS** | 0.22 (-0.14 - 0.58) | 33.3 | 0.61 (0.38 - 0.84) | 38.9 | 0.34 (-0.14 - 0.81) | 72.2 |
| **DLL** | 0.53 (0.30 - 0.77) | 38.9 | 0.72 (0.37 - 1.00) | 77.8 | 0.09 (-0.23 - 0.05) | 77.8 |
| **DLR** | 0.67 (0.45 - 0.89) | 44.4 | 0.86 (0.72 - 1.00) | 88.9 | 0.09 (-0.33 - 0.51) | 77.8 |
| **SLRDLL** | 0.14 (-0.25 - 0.52) | 50.0 | 0.32 (0.03 - 0.60) | 44.4 | 0.16 (-0.34 - 0.66) | 50.0 |
| **SLRDLR** | 0.51 (0.15 - 0.88) | 50.0 | 0.35 (-0.02 - 0.72) | 50.0 | 0.32 (0.05 - 0.60) | 33.3 |
| **PU** | 0.67 (0.44 - 0.90) | 44.4 | 0.72 (0.46 - 0.98) | 72.2 | 0.64 (0.00 - 1.00) | 94.4 |
| **CU** | 0.87 (0.75 - 0.98) | 55.6 | 0.74 (0.52 - 0.96) | 61.1 | 0.87 (0.75 - 0.99) | 66.7 |
| **Total Score ($ICC_{2,1}$)** | 0.50 (0.08 - 0.77) | 5.6 | 0.61 (0.21 - 0.84) | 16.7 | 0.58 (0.18 - 0.82) | 22.2 |

OHS = overhead squat; DLL = double lunge left; DLR = double lunge right; SLRDLL = single leg Romanian deadlift left; SLRDLR = single leg Romanian deadlift right; CU = chin-up; PU = push-up

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

**Table 4**: Video-based intra-rater agreement of key assessment areas

| | | EXP | | $NOV_6$ | | $NOV_7$ | | |
|---|---|---|---|---|---|---|---|---|
| | | $k_w$ (95% CI) | %Agree. | $k_w$ (95% CI) | %Agree. | $k_w$ (95% CI) | %Agree. | $k_w$ (95% |
| | Knee Alignment | 0.40 (-0.01 - 0.81) | 77.8 | 0.00 (0.00 - 0.00) | 94.4 | 0.64 (0.00 - 1.00) | 94.4 | 0.14 (-0.27 - 0. |
| **OHS** | Upper Quadrant | 0.77 (0.48 - 1.00) | 88.9 | 0.00 (0.00 - 0.00) | 94.4 | 0.44 (-0.21 - 1.00) | 88.9 | 0.59 (0.25 - 0. |
| | Triple Flexion | 1.00 (1.00 - 1.00) | 100.0 | 0.64 (0.27 - 1.00) | 83.3 | 0.40 (-0.11 - 0.91) | 83.3 | 0.01 (-0.52 - 0. |
| | Hip, Knee, Ankle | 0.63 (0.32 - 0.95) | 72.2 | 0.76 (0.47 - 1.00) | 83.3 | 0.55 (0.26 - 0.85) | 72.2 | 0.00 (-1.00 - -1 |
| **DLL** | Take Off Control | 0.78 (0.49 - 1.00) | 88.9 | 0.29 (-0.11 - 0.69) | 66.7 | 0.38 (-0.08 - 0.84) | 83.3 | 0.77 (0.48 - 1. |
| | Hip Range | na | 100.0 | 0.64 (0.00 - 1.00) | 94.4 | 0.78 (0.36 - 1.00) | 94.4 | 0.61 (0.14 - 1. |
| | Hip, Knee, Ankle | 0.54 (0.24 - 0.84) | 72.2 | 0.65 (0.34 - 0.96) | 77.8 | 0.44 (0.03 - 0.86) | 72.2 | 0.33 (-0.11 - 0. |
| **DLR** | Take Off Control | 0.80 (0.59 - 1.00) | 83.3 | 0.79 (0.55 - 1.00) | 83.3 | 0.67 (0.50 - 0.84) | 72.2 | 0.87 (0.68 - 1. |
| | Hip Range | 0.64 (0.00 - 1.00) | 94.4 | 0.93 (0.78 - 1.00) | 94.4 | 0.78 (0.36 - 1.00) | 88.9 | 0.94 (0.84 - 1. |
| | Single Leg Control | 0.77 (0.60 - 0.94) | 83.3 | -0.14 (-0.54 - 0.27) | 38.9 | 0.49 (0.17 - 0.82) | 72.2 | 0.35 (-0.10 - 0. |
| **SLRDLL** | Hip Control (S) | 0.75 (0.44 - 1.00) | 88.9 | na | 100.0 | 0.78 (0.49 - 1.00) | 88.9 | 0.39 (-0.01 - 0. |
| | Hinge Range | 0.23 (-0.35 - 0.80) | 72.2 | 0.82 (0.63 - 1.00) | 83.3 | 0.48 (0.12 - 0.84) | 72.2 | 0.43 (0.02 - 0. |
| | Single Leg Control | 0.79 (0.60 - 0.98) | 77.8 | 0.63 (0.37 - 0.89) | 61.1 | 0.64 (0.41 - 0.88) | 66.7 | 0.74 (0.47 - 1. |
| **SLRDLR** | Hip Control (S) | 0.63 | 72.2 | na | 100.0 | 0.15 | 50.0 | 0.47 |

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | (0.32 - 0.95) | | | | (-0.22 - 0.52) | | (0.08 - 0. |
| | Hinge Range | 0.82 (0.59 - 1.00) | 83.3 | 0.71 (0.45 - 0.98) | 77.8 | 0.45 (0.16 - 0.75) | 66.7 | 0.39 (-0.16 - 0. |
| | Total Body Control | 0.88 (0.65 - 1.00) | 94.4 | 0.43 (0.01 - 0.85) | 72.2 | 0.67 (0.36 - 0.99) | 83.3 | 0.25 (-0.27 - 0. |
| **PU** | Upper Quadrant | 0.67 (0.43 - 0.90) | 66.7 | 0.35 (-0.07 - 0.77) | 61.1 | 0.45 (0.11 - 0.79) | 66.7 | 0.60 (0.31 - 0. |
| | X30 Reps | 0.82 (0.59 - 1.00) | 83.3 | 0.57 (0.10 - 1.00) | 72.2 | 1.00 (1.00 - 1.00) | 100.0 | 0.71 (0.44 - 0. |
| | Scapular Rhythm | 0.45 (0.12 - 0.79) | 66.7 | 0.39 (0.00 - 0.77) | 55.6 | 0.67 (0.37 - 0.96) | 83.3 | 0.66 (0.39 - 0. |
| **CU** | Total Body Control | 0.69 (0.33 - 1.00) | 83.3 | 0.35 (-0.12 - 0.81) | 77.8 | 0.54 (0.15 - 0.93) | 83.3 | 0.49 (-0.09 - 1. |
| | X10 Reps | 1.00 (1.00 - 1.00) | 100.0 | 0.73 (0.42 - 1.00) | 83.3 | 1.00 (1.00 - 1.00) | 100.0 | 0.76 (0.49 - 1. |

OHS = overhead squat; DLL = double lunge left; DLR = double lunge right; SLRDLL = single leg Romanian deadlift
deadlift right; CU = chin-up; PU = push-up; S = sagittal axis

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

**Table 5**: Video-based intra-rater agreement of component movements and total score

| | EXP | | $NOV_6$ | | $NOV_7$ | | NO |
|---|---|---|---|---|---|---|---|
| | $k_w$ (95% CI) | %Agree. | $k_w$ (95% CI) | %Agree. | $k_w$ (95% CI) | %Agree. | $k_w$ (95% CI) |
| **OHS** | 0.84 (0.65 - 1.00) | 77.8 | 0.71 (0.44 - 0.98) | 83.3 | 0.52 (-0.01 - 1.00) | 77.8 | 0.29 (-0.13 - 0.70) |
| **DLL** | 0.75 (0.60 - 0.90) | 61.1 | 0.65 (0.27 - 1.00) | 55.6 | 0.74 (0.49 - 0.98) | 66.7 | 0.49 (0.07 - 0.91) |
| **DLR** | 0.90 (0.82 - 0.98) | 72.2 | 0.79 (0.58 - 1.00) | 66.7 | 0.75 (0.60 - 0.90) | 38.9 | 0.80 (0.63 - 0.98) |
| **SLRDLL** | 0.38 (0.00 - 0.75) | 61.1 | 0.41 (-0.01 - 0.84) | 44.4 | 0.36 (-0.07 - 0.79) | 55.6 | 0.42 (0.09 - 0.75) |
| **SLRDLR** | 0.78 (0.56 - 1.00) | 50.0 | 0.67 (0.48 - 0.85) | 50.0 | 0.41 (0.07 - 0.75) | 44.4 | 0.49 (0.19 - 0.79) |
| **PU** | 0.77 (0.57 - 0.98) | 66.7 | 0.51 (0.17 - 0.85) | 44.4 | 0.79 (0.60 - 0.99) | 66.7 | 0.63 (0.38 - 0.88) |
| **CU** | 0.91 (0.83 - 0.98) | 55.6 | 0.76 (0.53 - 0.89) | 44.4 | 0.92 (0.84 - 1.00) | 77.8 | 0.74 (0.53 - 0.95) |
| **Total Score ($ICC_{2,1}$)** | 0.93 (0.82 - 0.97) | 27.8 | 0.73 (0.41 - 0.89) | 16.7 | 0.69 (0.34 - 0.87) | 27.8 | 0.60 (0.22 - 0.83) |

OHS = overhead squat; DLL = double lunge left; DLR = double lunge right; SLRDLL = single leg Romanian deadlift
deadlift right; CU = chin-up; PU = push-up

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

**Table 6**: Typical error and minimum detectable change scores for component movements and total score when assessed in real-time

| | EXP | | $NOV_1$ | | $NOV_2$ | |
|---|---|---|---|---|---|---|
| | **TE (95% CI)** | **MDC** | **TE (95% CI)** | **MDC** | **TE (95% CI)** | **MDC** |
| **OHS** | 1.41 (1.06 - 2.12) | 3.91 | 0.94 (0.70 - 1.41) | 2.61 | 1.33 (1.00 - 2.00) | 3.69 |
| **DLL** | 0.71 (0.53 - 1.06) | 1.97 | 0.34 (0.26 - 0.51) | 0.94 | 0.33 (0.25 - 0.50) | 0.91 |
| **DLR** | 0.69 (0.51 - 1.03) | 1.91 | 0.23 (0.17 - 0.34) | 0.64 | 0.45 (0.34 - 0.68) | 1.25 |
| **SLRDLL** | 0.88 (0.66 - 1.32) | 2.44 | 0.54 (0.40 - 0.80) | 1.50 | 0.65 (0.49 - 0.98) | 1.80 |
| **SLRDLR** | 0.82 (0.62 - 1.23) | 2.27 | 0.66 (0.50 - 0.99) | 1.83 | 0.81 (0.61 - 1.22) | 2.25 |
| **PU** | 0.70 (0.52 - 1.04) | 1.94 | 0.36 (0.27 - 0.55) | 1.00 | 0.33 (0.25 - 0.50) | 0.91 |
| **CU** | 0.62 (0.47 - 0.93) | 1.72 | 0.58 (0.44 - 0.88) | 1.61 | 0.50 (0.38 - 0.75) | 1.39 |
| **Total Score** | 2.91 (2.18 - 4.36) | 8.07 | 1.91 (1.44 - 2.87) | 5.29 | 2.12 (1.59 - 3.18) | 5.88 |

OHS = overhead squat; DLL = double lunge left; DLR = double lunge right; SLRDLL = single leg Romanian deadlift left; SLRDLR = single leg Romanian deadlift right; CU = chin-up; PU = push-up

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

**Table 7**: Typical error and minimum detectable change scores for component movements and total score when assesse

| | EXP | | $NOV_6$ | | $NOV_7$ | | NOV |
|---|---|---|---|---|---|---|---|
| | **TE (95% CI)** | **MDC** | **TE (95% CI)** | **MDC** | **TE (95% CI)** | **MDC** | **TE (95% CI)** |
| **OHS** | 0.70 (0.52 - 1.04) | 1.94 | 0.29 (0.22 - 0.44) | 0.80 | 0.80 (0.60 - 1.20) | 2.22 | 1.41 (1.06 - 2.12 |
| **DLL** | 0.45 (0.34 - 0.68) | 1.25 | 0.57 (0.43 - 0.58) | 1.58 | 0.64 (0.48 - 0.95) | 1.77 | 0.73 (0.55 - 1.09 |
| **DLR** | 0.33 (0.24 - 0.49) | 0.91 | 0.64 (0.48 - 0.95) | 1.77 | 0.68 (0.51 - 1.02) | 1.88 | 0.49 (0.36 - 0.73 |
| **SLRDLL** | 0.88 (0.66 - 1.32) | 2.44 | 0.67 (0.50 - 1.00) | 1.86 | 0.82 (0.62 - 1.24) | 2.27 | 1.01 (0.76 - 1.51 |
| **SLRDLR** | 0.75 (0.56 - 1.12) | 2.08 | 0.43 (0.32 - 0.64) | 1.19 | 0.95 (0.71 - 1.42) | 2.63 | 0.79 (0.59 - 1.18 |
| **PU** | 0.69 (0.51 - 1.03) | 1.91 | 0.98 (0.74 - 1.48) | 2.72 | 0.50 (0.38 - 0.75) | 1.39 | 0.83 (0.63 - 1.25 |
| **CU** | 0.56 (0.42 - 0.83) | 1.55 | 0.78 (0.58 - 1.16) | 2.16 | 0.49 (0.36 - 0.73) | 1.36 | 0.72 (0.54 - 1.08 |
| **Total Score** | 1.08 (0.81 - 1.63) | 2.99 | 1.96 (1.47 - 2.93) | 5.43 | 2.55 (1.91 - 3.82) | 7.07 | 2.83 (2.13 - 4.25 |

OHS = overhead squat; DLL = double lunge left; DLR = double lunge right; SLRDLL = single leg Romanian deadlift
deadlift right; CU = chin-up; PU = push-up

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

**Table 8:** Inter-rater reliability of key assessment areas

| | | $ICC_{2,1}$ (95% CI) | |
|---|---|---|---|
| | | **Real-time** | **Video-based** |
| | Knee Alignment | 0.21 (0.06 - 0.45) | -0.05 (-0.13 - 0.12) |
| **OHS** | Upper Quadrant | 0.28 (0.12 - 0.53) | 0.19 (0.04 - 0.44) |
| | Triple Flexion | 0.20 (0.06 - 0.45) | 0.21 (0.06 - 0.46) |
| | Hip, Knee, Ankle | 0.30 (0.13 - 0.55) | 0.29 (0.10 - 0.55) |
| **DLL** | Take Off Control | 0.41 (0.22 - 0.65) | 0.40 (0.20 - 0.65) |
| | Hip Range | 0.31 (0.13 - 0.55) | 0.45 (0.24 - 0.68) |
| | Hip, Knee, Ankle | 0.09 (0.00 - 0.28) | 0.30 (0.12 - 0.56) |
| **DLR** | Take Off Control | 0.38 (0.19 - 0.62) | 0.66 (0.48 - 0.83) |
| | Hip Range | 0.23 (0.07 - 0.47) | 0.55 (0.35 - 0.76) |
| | Single Leg Control | 0.13 (0.00 - 0.36) | 0.21 (0.04 - 0.47) |
| **SLRDLL** | Hip Control (S) | 0.30 (0.13 - 0.55) | 0.10 (-0.03 - 0.32) |
| | Hinge Range | 0.23 (0.07 - 0.47) | 0.28 (0.10 - 0.53) |
| | Single Leg Control | 0.46 (0.27 - 0.69) | 0.46 (0.25 - 0.69) |
| **SLRDLR** | Hip Control (S) | 0.21 (0.07 - 0.45) | 0.12 (-0.01 - 0.36) |
| | Hinge Range | 0.31 (0.13 - 0.55) | 0.21 (0.05 - 0.46) |
| | Total Body Control | 0.30 (0.13 - 0.55) | 0.33 (0.14 - 0.58) |
| **PU** | Upper Quadrant | 0.15 (0.03 - 0.36) | 0.26 (0.08 - 0.52) |
| | X30 Reps | 1.00 | 0.63 (0.49 - 0.84) |
| | Scapular Rhythm | 0.12 (0.01 - 0.33) | 0.15 (0.01 - 0.39) |
| **CU** | Total Body Control | 0.41 (0.22 - 0.65) | 0.57 (0.36 - 0.77) |
| | X10 Reps | 0.96 (0.93 - 0.98) | 0.78 (0.62 - 0.90) |

OHS = overhead squat; DLL = double lunge left; DLR = double lunge right; SLRDLL = single leg Romanian deadlift left; SLRDLR = single leg Romanian deadlift right; CU = chin-up; PU = push-up; S = sagittal axis

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

**Table 9:** Inter-rater reliability of component movements and total score

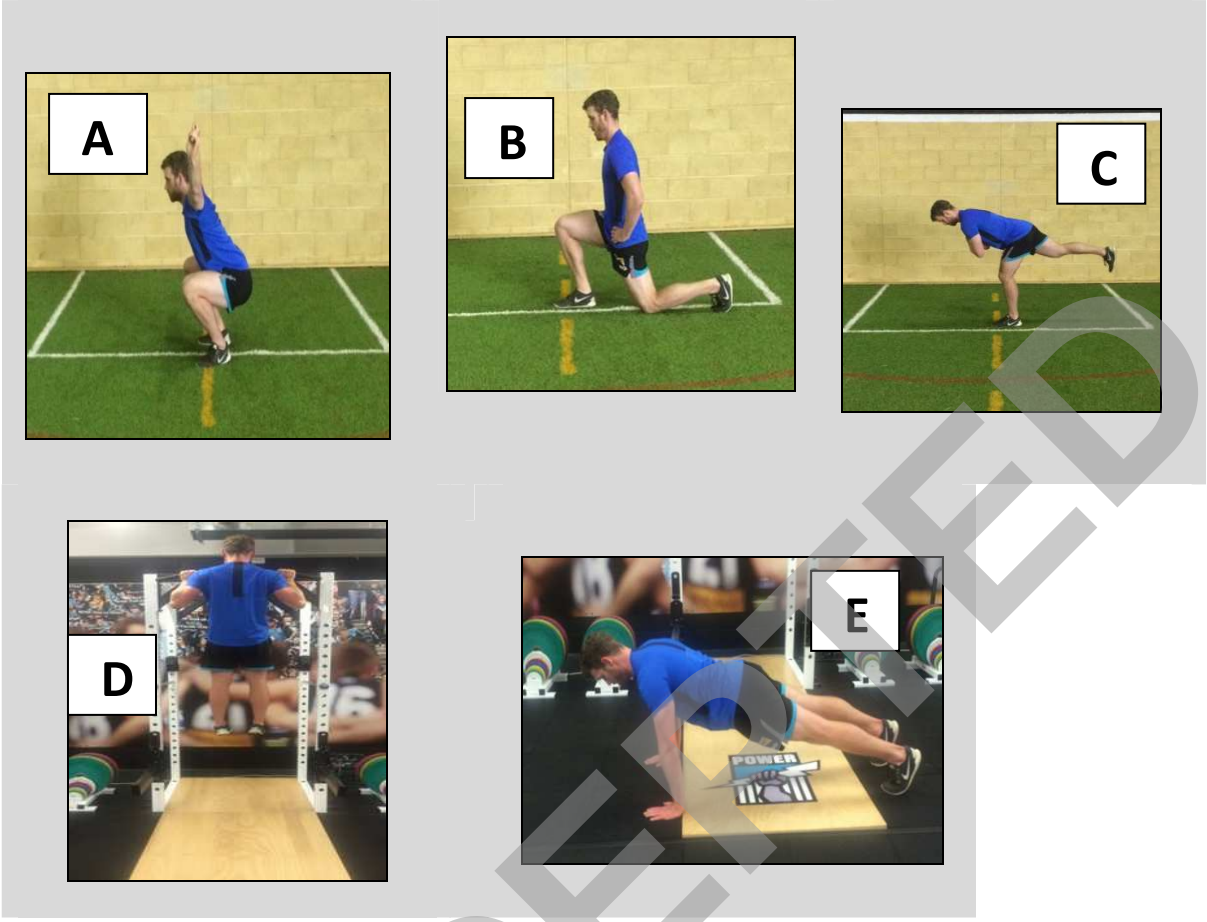
| | $ICC_{2,1}$ (95% CI) | |
|---|---|---|
| | **Real-time** | **Video-based** |
| **OHS** | 0.26 (0.10 - 0.51) | 0.16 (0.03 - 0.40) |
| **DLL** | 0.54 (0.33 - 0.75) | 0.54 (0.34 - 0.75) |
| **DLR** | 0.38 (0.19 - 0.63) | 0.67 (0.48 - 0.84) |
| **SLRDLL** | 0.14 (0.02 - 0.36) | 0.19 (0.04 - 0.44) |
| **SLRDLR** | 0.36 (0.18 - 0.60) | 0.33 (0.14 - 0.59) |
| **PU** | 0.63 (0.41 - 0.81) | 0.53 (0.32 - 0.75) |
| **CU** | 0.83 (0.69 - 0.93) | 0.68 (0.48 - 0.84) |
| **Total Score** | 0.52 (0.28 - 0.75) | 0.47 (0.24 - 0.71) |

OHS = overhead squat; DLL = double lunge left; DLR = double lunge right; SLRDLL = single leg Romanian deadlift left; SLRDLR = single leg Romanian deadlift right; CU = chin-up; PU = push-up

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association

**Figure 1**: The Five component movements of the AAA




A = overhead squat; B = double lunge; C = single leg Romanian deadlift; D = chin-up; E = push-up

ACCEPTED

Copyright © 2017 National Strength and Conditioning Association