The Intra- and Inter-rater Reliability of an Arm Care Screening Tool in High School Baseball Coaches

Kyle A. Matsel1, Stephanie N. Brown2, Matt C. Hoch2, Robert J. Butler3, Philip M. Westgate1, Terry R. Malone2, Tim L. Uhl†

1 University of Kentucky; University of Evansville, 2 University of Kentucky, 3 Saint Louis Cardinals Baseball Organization

Keywords: screening, reliability, movement system, baseball

https://doi.org/10.26603/001c.29680

Background
Preseason movement screening can identify modifiable risk factors, deterioration of function, and potential for injury in baseball players. Limited resources and time intensive testing procedures prevent high school coaches from accurately performing frequent movement screens on their players.

Purpose
To establish the intra-rater and inter-rater reliability of a novel arm care screening tool based on the concepts of the Functional Movement Screen (FMS™) and Selective Functional Movement Assessment (SFMA™) in high school coaches.

Study Design
Methodological intra- and inter-rater reliability study

Methods
Thirty-one male high school baseball players (15.9 years ± 1.06) were independently scored on the Arm Care Screen (ACS) by three examiners (two coaches, one physical therapist) in real-time and again seven days later by reviewing video recordings of each players' initial screening performance. Results from each examiner were compared within and between raters using Cohen’s kappa and percent absolute agreement.

Results
Substantial to excellent intra-rater and inter-rater reliability were established among all raters for each component of the ACS. The mean Cohen’s kappa coefficient for intra-rater reliability was 0.76 (95% confidence interval, 0.54-0.95) and percent absolute agreement ranged from 0.82-0.94 among all raters. Inter-rater reliability demonstrated a mean Cohen’s kappa value of 0.89 (95% confidence interval, 0.77-0.99) while percent absolute agreement between raters ranged from 0.81-1.00. Intra- and inter-rater reliability did not differ between raters with various movement screening experience (p>0.05).

Conclusions
High school baseball coaches with limited experience screening movement can reliably score all three components of the ACS in less than three minutes with minimal training.

Corresponding Author:
Kyle A Matsel
University of Evansville
515 Bob Jones Way
Evansville, IN 47708
Phone: (812) 488-5105
E-mail: km218@evansville.edu
Level of Evidence
Level 3, Reliability study

INTRODUCTION

Baseball players are susceptible to shoulder and elbow injuries which can result in missed time from sport.1 Microtrauma from repetitive physical misuse of the kinetic chain is a common avenue for development of injury.2,3 The overhead throwing motion, common in baseball, requires contribution of the entire body, specifically the hips, torso, and shoulder girdle.2 Limitations in any of these body regions can result in large amounts of mechanical stress placed on the upper extremity.3,4 Despite well-established pitching volume guidelines, injury rates in high school baseball players have continued to increase.5 Over the past two decades, elbow injuries, specifically ulnar collateral ligament reconstruction, (UCLR) have become more frequent in adolescents and are expected to continue to rise6,7 As a result, greater attention to identification and management of intrinsic modifiable risk factors such as strength, flexibility, and neuromuscular control is warranted for more comprehensive injury management.8

Preseason examinations have identified modifiable risk factors associated with increased injury rates during the season.9,10 Movement screening can provide useful information specific to deterioration of function and potential for injury. Two movement examination tools that exist are the Functional Movement Screen (FMS™) and the Selective Functional Movement Assessment (SFMA) which are designed to identify major limitations and/or asymmetries which could contribute to musculoskeletal pain or movement deficits.11–15 In high school baseball players, Lee et al.14 reported that FMS™ composite scores and individual test performance declined over the course of the season. Most studies have focused on singular risk factors in common body regions while overlooking the presence of additional impairments throughout the kinetic chain.14 In collegiate baseball players, Busch et al.15 reported that poor FMS™ and SFMA shoulder mobility patterns were associated with five and six-fold increased odds of having an overuse injury during preseason training, respectively.15 However, Busch et al.15 limited screening to only upper quarter tests and failed to include additional lower quarter movement patterns which would be more representative of the entire kinetic chain.

Physical limitations throughout the kinetic chain can contribute to upper extremity overuse injury in baseball players.2,16 The functional relationship between the upper extremities, spine, and hips required for kinetic linkage in rotational athletes warrants a more comprehensive screen of the entire body.17 Although the shoulder and elbow are common areas for symptom development, increased physiological stress on these joints can be produced by other remote regions in the body.18,19 Risk factors such as glenohumeral internal rotation deficit (GIRD),20–24 limited hip internal rotation ROM,25–28 limited hip external rotation,29 thoracic spine mobility,30 total shoulder ROM,22–24,31–33 and dynamic single leg balance34 have been identified during comprehensive pre-season examinations and can result in time loss from sport.

In most circumstances, high school coaches have limited resources needed to perform frequent movement screenings on their players. External factors such as funding for training,35 time constraints36 and insufficient staffing limit implementation of injury screening. Specifically, the FMS™ requires testing equipment, multiple hours of training for certification, and takes approximately 10–15 minutes to perform on each player.11 The SFMA is a clinical assessment designed for practitioners with a license to diagnose and treat pain37 and therefore is out of the scope of practice for the coach. Thus, there is a need for a field expedient screening tool which can measure physical function throughout the kinetic chain that can quickly be administered by coaches. Frequent screening throughout the season could identify impairments and inform arm care training programs to improve strength and flexibility.38

To date there are no studies which have explored a field expedient screening tool that can be quickly performed by high school baseball coaches. The primary purpose of this study was to establish the intra-rater and inter-rater reliability of a novel arm care screening tool based on the concepts of the FMS™ and SFMA in high school coaches. It was hypothesized that the Arm Care Screen (ACS) will demonstrate moderate intra-rater and inter-rater reliability with a Cohen’s kappa value >0.40 when administered by high school baseball coaches.

MATERIALS AND METHODS

STUDY DESIGN

A prospective methodological cohort design was used to establish the intra-rater and inter-rater reliability of the ACS among high school baseball coaches. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement was used for quality reporting.39 Approval was granted from the institutional review board at the University of Kentucky and informed consent and assent forms were obtained prior to data collection.

PARTICIPANTS

A minimum sample size of 24 players was needed to achieve a Cohen’s kappa value of 0.40 with an alpha of 0.05 and 80% power for a two-tailed test. Anticipating 10% of the participants would have missing or incomplete data the planned target sample size was 27 participants. A convenience sample of 31 male high school baseball players from a single local team volunteered to participate in this study. The head coach allowed the researchers to attend team workouts to recruit and test the players. Inclusion criteria required the participant to be a current and active member of a men’s varsity, junior varsity, or freshman high school baseball team. Exclusion criteria included the inability to participate in sport due to current injury, recent surgery (within three months), physician restriction, recent concussion (within one month) or vestibular issue. Players who were under the
age of 14 or older than 18 were excluded (Figure 1).

PROCEDURES

Data collection occurred over a four-week period in the fall offseason during team workouts. Physical testing occurred during a single session at the participants’ local high school baseball field. Participants completed a demographic questionnaire which included information about their age, height, weight, playing position, baseball experience, and current injury status. All participants were scored once on ACS performance in real-time while being videotaped for later assessment of intra-rater reliability.

The ACS is a modified movement-based screening tool which utilizes components of the FMS™, SFMA, and Y-Balance Test-Lower Quarter developed to improve field expediency and reduce scoring complexity. The ACS consists of three tests including, 1) reciprocal shoulder mobility, 2) 90/90 total body rotation, and 3) lower body (LB) diagonal reach (Figure 2). Each component of the ACS was scored as pass or fail per the criteria below on both the right and left sides. Pain with testing was recorded but did not factor into the scoring criteria.

Reciprocal Shoulder Mobility: The participant began in standing with feet together and both hands open. The participant simultaneously reached one hand behind their head and other hand behind and up their back, similar to an Apley’s Scratch test position, assuming an extended and internally rotated position with one shoulder and a flexed and externally rotated position with the other. The arms must move in one smooth motion and tall posture must be maintained while the participant attempts to touch the fingertips of both hands together. Inability to touch right and left fingertips together on both reach directions was considered a failure in the test.

90/90 Total Body Rotation: The participant assumes a standing position with feet together, toes pointing forward and arms in the 90/90 position (90° shoulder abduction and 90° elbow flexion). The participant rotates their entire body including the hips, shoulders, and head as far as possible to the right while the foot position remained unchanged. Inability to see the back shoulder when viewing the participant from behind on both sides was considered failure on the test.

Lower Body Diagonal Reach: The participant stood two shoe lengths away from a wall and while maintaining single leg balance on one foot the participant reaches with the opposite foot behind and across their body to try to touch the point on the wall just above the ground five consecutive times without the foot touching down or losing balance. Inability to touch the wall five consecutive times or loss of balance on either side was considered failure on the test.

INTRA-RATER AND INTER-RATER RELIABILITY

Two high school baseball coaches were recruited to each score the ACS at two different time points. Prior to data collection, both coaches signed the informed consent and underwent a one-hour electronic ACS training session performed by the lead author (K AM). Following the training session, both coaches were required to pass an online ACS competency examination with a score of ≥80% prior to data collection. Additional online training and a retake exam would be provided for coaches scoring <80% on the exam. Remediation was not required in this study since both coaches passed their exam on the first attempt.

The participants were given the ACS instructions by one rater (K AM) while the testing was observed by two other raters simultaneously in real-time. All raters were blinded to each other’s scoring. The PT rater (K AM) was a physical therapist with 10 years of experience working with baseball players and coach rater 1 (C1) (JB) and coach rater 2 (C2) (TO) were assistant high school baseball coaches with greater than five years of coaching experience. All participants were video recorded performing the ACS testing procedures during the live testing using an iPhone 10Xr cell phone placed approximately 15 feet directly posterior to the participant. Following a seven-day washout period, the same three raters rescoring the ACS performance using the
video recordings taken during the live testing and were blinded to each other's results. To minimize recall bias, the order of the videos was randomized to differ from the original live testing order.

STATISTICAL METHODS

Descriptive statistics including means and standard deviations (SD) or frequency counts were calculated as appropriate. Intra-rater and inter-rater reliability for the categorical scores of each component of the ACS were compared within and between (PT-C1, PT-C2, and C1-C2 each rater using Cohen's kappa coefficient with 95% confidence intervals (CI95%) and percent absolute agreement. The Cohen's kappa coefficient quantifies the strength of agreement and was interpreted as: ≤0.40 = poor to slight, 0.41-0.60 = moderate, 0.61-0.80 = substantial, ≥0.80 = excellent. All data analyses were performed with SPSS statistical software (IBM SPSS Statistics for Mac, Version 27.0). An alpha level of p < 0.05 was considered statistically significant for all tests.

RESULTS

Demographic characteristics of all participants are provided in Table 1. The mean age ± SD of the participants in this sample was 15.9 ± 1.06, 45.2% (n=14/31) were high school sophomores, and 25.8% (n=8/31) were primarily pitchers. A total of 29% (n=9/31) of participants reported pain, which was defined as discomfort beyond just a stretch or soreness, with at least one component of the ACS, however, pain was not part of the passing criteria for the tests and did not affect the scoring. All 31 participants completed the testing procedures, and the results were used for data analysis.

INTRA-RATER AND INTER-RATER RELIABILITY

Results of intra-rater reliability scores after a seven-day washout period within all three raters are presented in Table 2 including corresponding Cohen's kappa values with 95% confidence intervals and percent agreement. Cohen's kappa values for the three component tests of the ACS (scored pass or fail) demonstrated moderate to excellent intra-rater agreement. The mean intra-rater reliability for all raters was substantial with a Cohen's kappa value of 0.76 (95% CI, 0.54-0.95) and a mean absolute agreement was 89%. The PT rater demonstrated higher intra-rater reliability compared to C1 and C2, but these differences were not statistically significant (p > 0.05).

The results for ACS same day inter-rater reliability with Cohen's kappa values, 95% confidences, and percent absolute agreement are presented in Table 3. Cohen kappa values ranged from moderate to excellent agreement between all three raters depending on the specific movement component of the ACS. The mean inter-rater reliability for all raters was excellent with a Cohen's kappa value of 0.89 (95% CI, 0.77-0.99) and a mean absolute agreement was 95%. The overall mean kappa agreement when comparing each raters' performance on all ACS components demonstrated near perfect agreement and did not differ significantly between raters (p > 0.05).

DISCUSSION

The purpose of this study was to establish the intra- and inter-rater reliability of a field expedient arm care screening tool in a sample of high school baseball players. The findings from this study supports the primary hypothesis that
Table 2. Intra-rater Reliability of ACS Components (n=31)

<table>
<thead>
<tr>
<th></th>
<th>PT Rater</th>
<th>C1 Rater</th>
<th>C2 Rater</th>
<th>All Raters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kappa (CI95%)</td>
<td>% Agree</td>
<td>Kappa (CI95%)</td>
<td>% Agree</td>
</tr>
<tr>
<td>NonDom Shoulder Mobility</td>
<td>0.81 (0.60-1.0)</td>
<td>0.90</td>
<td>0.62 (0.36-0.87)</td>
<td>0.81</td>
</tr>
<tr>
<td>Dom Shoulder Mobility</td>
<td>1.0 (1.0-1.0)</td>
<td>1.00</td>
<td>0.78 (0.56-1.0)</td>
<td>0.90</td>
</tr>
<tr>
<td>NonDom Total Body Rotation</td>
<td>0.80 (0.59-1.0)</td>
<td>0.90</td>
<td>0.74 (0.49-0.98)</td>
<td>0.87</td>
</tr>
<tr>
<td>Dom Total Body Rotation</td>
<td>0.83 (0.61-1.0)</td>
<td>0.94</td>
<td>0.74 (0.46-1.0)</td>
<td>0.90</td>
</tr>
<tr>
<td>Stride LB Diagonal Reach</td>
<td>0.60 (0.32-0.87)</td>
<td>0.81</td>
<td>0.87 (0.70-0.78)</td>
<td>0.94</td>
</tr>
<tr>
<td>Stance LB Diagonal Reach</td>
<td>0.68 (0.43-0.93)</td>
<td>0.84</td>
<td>0.50 (0.22-0.83)</td>
<td>0.74</td>
</tr>
<tr>
<td>Mean</td>
<td>0.79 (0.59-0.97)</td>
<td>0.90</td>
<td>0.71 (0.47-0.91)</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Nondom=nondominant arm, Dom=dominant arm, LB=lower body, CI95%=95% confidence interval

Table 3. Inter-rater Reliability of ACS Components (n=31)

<table>
<thead>
<tr>
<th></th>
<th>PT Rater vs. C1</th>
<th>PT Rater vs. C2</th>
<th>C1 Rater vs. C2</th>
<th>All Raters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kappa (CI95%)</td>
<td>% Agree</td>
<td>Kappa (CI95%)</td>
<td>% Agree</td>
</tr>
<tr>
<td>NonDom Shoulder Mobility</td>
<td>1.0 (1.0-1.0)</td>
<td>1.00</td>
<td>1.0 (1.0-1.0)</td>
<td>1.00</td>
</tr>
<tr>
<td>Dom Shoulder Mobility</td>
<td>0.92 (0.76-1.0)</td>
<td>0.97</td>
<td>0.92 (0.76-1.0)</td>
<td>0.97</td>
</tr>
<tr>
<td>NonDom Total Body Rotation</td>
<td>0.73 (0.48-0.97)</td>
<td>0.87</td>
<td>0.86 (0.67-1.0)</td>
<td>0.93</td>
</tr>
<tr>
<td>Dom Total Body Rotation</td>
<td>0.91 (0.74-1.0)</td>
<td>0.97</td>
<td>0.91 (0.75-1.0)</td>
<td>0.97</td>
</tr>
<tr>
<td>Stride LB Diagonal Reach</td>
<td>0.81 (0.60-1.0)</td>
<td>0.93</td>
<td>0.67 (0.41-0.93)</td>
<td>0.84</td>
</tr>
<tr>
<td>Stance LB Diagonal Reach</td>
<td>0.94 (0.81-1.0)</td>
<td>0.97</td>
<td>0.87 (0.70-1.0)</td>
<td>0.94</td>
</tr>
<tr>
<td>Mean</td>
<td>0.88 (0.73-0.99)</td>
<td>0.95</td>
<td>0.87 (0.72-0.99)</td>
<td>0.94</td>
</tr>
</tbody>
</table>

* Nondom=nondominant arm, Dom=dominant arm, LB=lower body, CI95%=95% confidence interval

baseball coaches can reliably administer the ACS to screen high school players for physical limitations which may contribute to time loss from sport and determine need for further evaluation by a healthcare professional. All three subcomponents of the ACS, scored as pass or fail, exhibited substantial to excellent intra- and inter-rater reliability among all three raters regardless of their coaching or movement screening experience. The reliability of movement-based screening tools has been evaluated in multiple populations among raters of differing professional backgrounds such as physical therapists, certified athletic trainers (ATC), and strength and conditioning specialists.41–43 The results of the current study are similar to previous literature investigating rater agree-
ment of the FMS™ and SFM. Mininck et al. reported substantial to excellent interrater reliability (k = 0.74-1.0) among expert FMS raters and novice raters while scoring FMS™ performance from video recordings. Likewise, Teyhen et al. reported moderate to excellent interrater and test-retest agreement in physical therapy students on the component tests of the FMS when scored in real time. Moderate intra-rater (ICC = 0.75; 95% CI, 0.53-0.87) reliability has been observed among ATCs and athletic training students when evaluating the composite score of the FMS™ in real time. Conversely, Shultz et al. which included strength and conditioning coaches, showed fair to poor inter-rater reliability (K = 0.38; 95% CI, 0.35-0.41) of FMSTM subsets.

The SFMA categorical scoring criteria have demonstrated slight to substantial intra-rater (k=0.48-0.85) and inter-rater (k=0.20-0.76) reliability in healthy adults. However, reliability conclusions are limited to only physical therapists and ATCs to date. This is the first study to specifically examine the ability of high school baseball coaches to screen movement patterns accurately and consistently. Previously, three studies exploring the reliability of the SFMA used video analysis to aid with rater scoring. The methodology of the current study included both real-time scoring and video analysis to establish reliability of the ACS. Although video analysis was consistent with previous SFMA reliability studies, the utility of the ACS is more applicable to coaches in scoring players in real-time during practice which warranted exploration of live screening accuracy.

Previous authors have reported that raters with more movement testing experience have better intra- and interrater reliability compared to less experienced raters. In the current study, there was no reliability differences between raters even though the PT rater had 10 years of experience screening movement while C1 and C2 only had a one-hour educational session prior to data collection. It is likely that the reduced number of categorical scoring options compared to other movement screens minimized errors among the raters. By dichotomizing each subset of the ACS as pass or fail scoring complexity is reduced and the raters are better able to agree on the testing results regardless of their experience.

Resources, experience, and staffing limitations prevent high school coaches from performing comprehensive musculoskeletal assessments and testing on their players multiple times during the season. The ACS provides coaches with a tool to track changes in physical function throughout the season so that exercise intervention and/or further evaluation can be recommended prior to the onset of injury. High school coaches can perform the ACS is less than three minutes which is more time efficient and feasible compared to the 10-15 minutes needed to administer the FMS. Furthermore, minimal training (~one hour) is required to be proficient in scoring the ACS with no additional costs or certification required to implement. While the ACS has shown to be reliable, validation of the screening tool is warranted prior to mainstream use in high school athletics. Future research should investigate the discriminability of the ACS to detect physical impairments and risk factors in baseball players.

LIMITATIONS

The authors acknowledge that this study is not without limitations. First, external validity of the intra-rater reliability may have been affected by performing live screens initially but rescreening seven days after from a video recording. During live screening the raters where able to view the participants movement from multiple different angles as opposed to only a singular posterior view on the video recording. Despite the different scoring approaches, Cohen's kappa values were not drastically affected as intra-rater reliability was excellent among all raters. Secondly, the current study focused specifically on high school baseball players who played a variety of positions. The results of this study may not be generalizable to baseball players at the professional and collegiate level. Only high school players were included to maximize homogeneity of the sample, however, both position players and pitchers were also included. Musculoskeletal function differs between pitchers and position players and likely resulted in increased heterogeneity among the participants.

CONCLUSION

High school baseball coaches with limited experience screening movement can reliably score all three components of the ACS in less than three minutes with minimal training. All raters demonstrated substantial to excellent intra- and inter-rater reliability which did not differ based on screening experience. Therefore, the ACS is a highly feasible movement screening option for implementation in the high school baseball environment. Future research should focus on exploring the discriminant validity of the ACS at identifying physical impairments and injury risk factors.

ACKNOWLEDGEMENTS

The authors would like to thank Evansville North High School and head baseball coach Jeremy Jones for access to the players and assistant coaches Tyler Owen and Jordan Bedwell for assistance with data collection. None of the authors of this article have any conflicts of interest or financial conflicts to report.

Submitted: June 11, 2021 CST, Accepted: September 04, 2021 CST
REFERENCES


