Note: this is the final draft of the article:

Rowsome K, Comerford M, Mottram S, Samuel D, Stokes M

Movement control testing of older people in community settings: description of a screening tool and intra-rater reliability *Working Pap Health Sci* 2016; 1:15 Spring: 1-12 ISSN 2051-6266 / 20150091

Accepted: 25th April 2016 for publication in *Working Papers in the Health Sciences*

Movement control testing of older people in community settings: description of a screening tool and intra-rater reliability

Keira Rowsome, MSc[1], MSc Physiotherapy Student

Mark Comerford, B.Phty[1,2], Visiting Research Fellow

Sarah Mottram, MSc[1,2,3], Visiting Research Fellow

Dinesh Samuel PhD[1], Lecturer in Physiotherapy

Maria Stokes, PhD[1,3], Professor of Musculoskeletal Rehabilitation

1. Faculty of Health Sciences, University of Southampton, UK
2. Movement Performance Solutions, Bristol, UK
3. Arthritis Research UK Centre for Sport, Exercise and Osteoarthritis, UK

RUNNING HEAD: Movement control tests for older people

Corresponding Author:
Professor Maria Stokes
Faculty of Health Sciences, Building 45
University of Southampton
Highfield Campus
Southampton
SO17 1BJ
Tel: +44(0)2380 596868
E-mail: M.Stokes@soton.ac.uk

Abstract

Objective: To determine the intra-rater reliability of a newly developed movement screening tool; the 'Movement control screen for older people in community settings'. The movement screening tool aims to identify movement control impairments which can potentially influence movement function. **Method:** Thirty one active female recreational golfers, aged 65-77 years, carried out three movement control tests included in the screening tool. Performance was video-recorded to enable repeated ratings. Each test was evaluated by criteria which were rated as pass or fail and ratings were carried out three weeks apart to examine intra-rater reliability. Reliability was assessed using percentage agreement and Cohen's Kappa. **Results:** Percentage agreement for each test ranged from 93.0-97.3%, with an overall mean agreement of 95.5%. Kappa values for test scores ranged from 0.35-0.90. Percentage agreement for individual criteria ranged from 83.0-100.0%, with kappa values ranging from 0.00-1.00. **Discussion:** Acceptable intra-rater reliability was established for overall tests scores of the screening tool but certain criteria were identified as being less reliable than others. Recommendations are made for refinement of some criteria to improve reliability of the screening tool.

Key words: Movement control, movement screening, reliability, impairments

Introduction

Functional ability in the performance of everyday tasks declines with age and is associated with declines in skeletal muscle mass (Ikezoe et al, 2010) and strength (Buford et al, 2012). For example, the ability to rise from a chair and ascend and descend stairs are important daily activities, which require sufficient lower extremity muscle strength and power (Samuel et al, 2011). The most prominent contributing factor to the decline in muscle mass, strength and function in older people is reported to be physical inactivity, as it induces atrophy (Morley, 2012; Wang and Bai, 2012). It is therefore important for older people to remain active in order to reduce the rate of decline in physical functioning and improve quality of life (Heesch et al, 2012).

Participation in golf has increased over the years and is particularly popular amongst the older population (Siegenthaler and O'Dell, 2003; Torres-Ronda et al, 2011). It has been suggested that golf has the potential to contribute to the concept of successful ageing (i.e. physical, mental and social well-being in older age) due its physical characteristics and cognitive effort (Siegenthaler and O'Dell, 2003). Research indicates that strength, flexibility and balance are important characteristics of golf (Lephard et al, 2007). Also, Lephart et al. (2007) reported significantly greater hip, torso and shoulder strength and flexibility, and balance, following a golf-specific training programme. Furthermore, golf requires good movement control, so a test of movement control i.e. quality of movement, would be beneficial for assessing this aspect of the effects of golf. It has been highlighted that movement screening is important to enable specific exercises to correct poor control of movement (McNeil, 2014).

Common performance measures include the Timed Up and Go test (Podsialdo and Richardson, 1991), the Berg Balance Scale (Berg et al, 1989) and the Performance-Orientated

Assessment of Mobility (Tinetti, 1986) but they do not assess control of movement. In addition, these existing performance measures are predominantly developed for frail and sedentary populations, hence there is a ceiling effect for more active people.

It is important to study active individuals and develop movement control measures suitable for a range of functional abilities. When active individuals present with conditions that are common with ageing, such as stroke, osteoarthritis and musculoskeletal injuries, treatment interventions should aim to enable those individuals to resume their preferred activities, where possible, and not simply aim for the same goals as less active people of the same age.

The 'Movement Control Screen for Older People in Community Settings' is a novel screening tool developed to assess quality of movement. The screening tool includes three tests, which have been adapted from The Performance Matrix, a tool comprising a large database of tests used to assess movement, primarily in sports and occupational populations, e.g. military. Tests included in The Performance Matrix have been described in the literature, such as 'the double knee swing test' (McNeil, 2014) and 'the single leg small knee bend + lunge-lean (Monnier et al, 2012). The Foundation Matrix is the entry level screen within the Performance Matrix database and is designed to identify performance related inefficient control of movement in the kinetic chain. The inter-rater and intra-rater reliability of the Foundation Matrix have been found to be acceptable, with excellent reliability for overall test scores and variable for individual tests (Mischiati et al., 2015).

The tests in the 'Movement control screen for older people in community settings' are designed to identify movement control impairments that adversely affect functional ability. They are intended to be challenging and replicate aspects of control of posture/movement usually required during functional tasks relevant to older people. Furthermore, the movement control tests are intended to be applicable across the full range of functional abilities so as to avoid the ceiling effect of many other functional performance measures.

Existing movement control tests have been evaluated for reliability, such as 'The Functional Movement Screen' (FMS) (Teyhen et al, 2012; Gribble et al, 2013), movement control tests from the Performance Matrix (Monnier et al, 2012), the Foundation Matrix (Mischiati et al, 2015) and movement control tests of the lumbar spine (Luomajoki et al, 2007). The 'Movement control screen for older people in community settings' is a newly developed movement control test and no reliability studies have been undertaken. It is important to test the reliability, as these observational tests are sensitive to the subjective interpretation by the rater, so poor reliability would indicate, for example, that the instructions for the rater in the assessment criteria may not be clear enough. Therefore, the aim of the present study was to establish the intra-rater reliability of this new screening tool in a group of moderately active older females.

Method

The study was approved by the Ethics Committee of the Faculty of Health Sciences, University of Southampton. All participants received a participant information sheet and provided written informed consent prior to participation.

Participants

A sample of convenience consisted of 31 active, older females, aged 65 to 77 years (mean ± 1 standard deviation = 69.2 ± 3.5). Participants were required to play at least one full round of golf (18 holes), by walking and not using a buggy, at least once a week. Exclusion criteria were: body mass index above 30kg/m^2; neurological conditions such as stroke, multiple sclerosis or Parkinson's disease; active rheumatoid or osteoarthritis restricting the ability to carry out functional tasks; severe injury or major operations resulting in immobility in the past 5 years; cognitive impairments; medically uncontrolled conditions, such as diabetes, hypertension or hypotension; receiving treatment for cancer; on medication affecting muscle function; or taking part in other research studies involving training muscles.

Forty-four volunteers were recruited through posters, websites and advertising (radio and ladies golf day at a county event). Seven potential participants dropped out prior to data collection (mainly due to unavailability during the testing time frame and not all gave a reason) and six were unsuitable to participate, as determined by a screening questionnaire.

Protocol

Participants were screened for eligibility over the phone. The screening questionnaire included questions from the Physical Activity Scale for the Elderly (PASE) a valid and reliable tool (Loland, 2002) in order to establish the participants' activity levels.

Participants attended a research laboratory in the Faculty of Health Sciences at the University of Southampton for one 90 minute session. The measures reported in the present paper formed part of a larger study that also assessed muscle characteristics (e.g. strength and size using ultrasound imaging) and functional tests (e.g. timed up and go, sit to stand; paper in preparation). Movement control screening was completed in 15 minutes. Participants agreed not to take part in strenuous activity 48 hours before their testing session. All participants wore shorts and t-shirts in order to allow for observation of movement.

Prior to testing, body mass and height were measured using a weighing scale and stadiometer, respectively.

Movement control tests

The 'Movement control screen for older people in community settings' consists of three tests (Table 1); Test 1: Sit to stand with arm lift (Figure 1a); Test 2: trunk lean with knee bend and opposite arm lift (Figure 1b); and Test 3: Chest rotation with neutral head and pelvis (Figure 1c). A video camera (Sony HDR-CX280E) was used to record the participants performing the tests. Test 1 was carried out twice so that it could be recorded from both anterior and lateral views.

Table 1: Movement Control Tests in order of performance. Description of tests that form the 'Movement control impairment screen for older people in community settings'. Criteria for evaluating performance of the test are described in Appendix 2.

Test	Title	Description
1*	Sit-to-stand + arm lift	Sitting: forward trunk lean approx. 30°, then stand keeping knees over feet. Alternate arm lift to 90° flexion & lower, maintaining neutral scapular position (Figure 1a)
2	Trunk lean with knee bend and opposite arm lift	With hand support on bench/table, forward trunk lean from hips approx. 45°. Single knee flexion and opposite arm lift to horizontal shoulder flexion (Figure 1b)
3	Chest rotation with neutral head and pelvis	Bilateral small knee bend with thoracic rotation and neutral head, pelvis & legs (Figure 1c)

*Test carried out twice to allow recording of anterior and lateral views

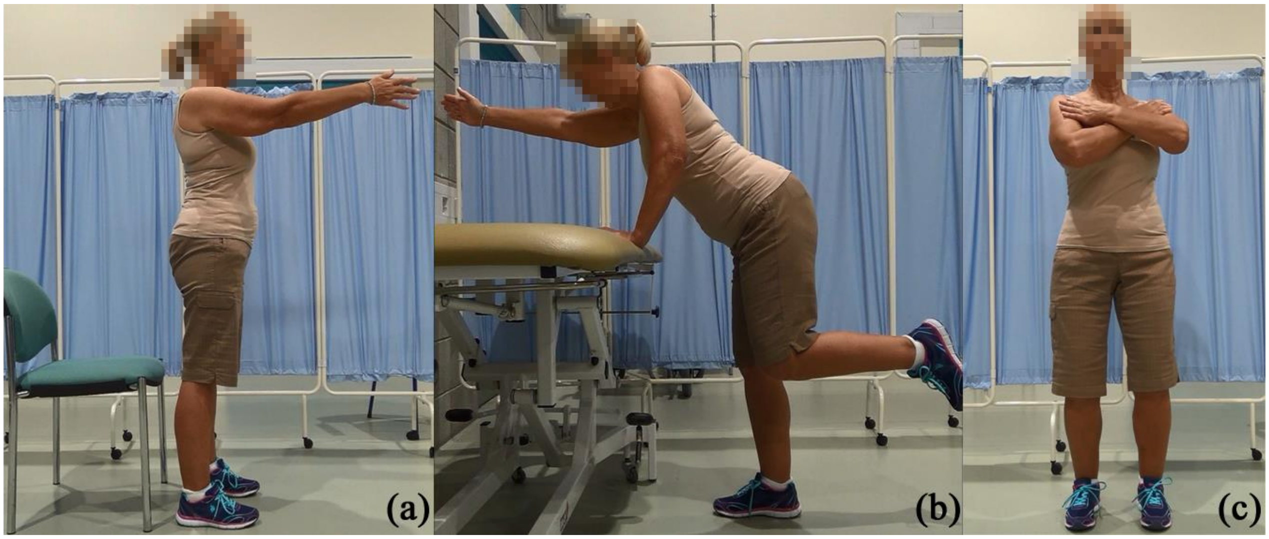

Figure 1: Participant carrying out movement control tests: a) Test 1: sit-to-stand + arm lift; b) Test 2: trunk lean with knee bend and opposite arm lift; c) Test 3: chest rotation with neutral head and pelvis (showing start of rotation to the participant's left side)

Prior to the testing procedure, each participant viewed videos of the correct performance for each test and the investigator verbally explained the tests in detail. Participants were also given the opportunity to practise each test a maximum of twice, so they became familiar with the movements and avoid the risk of fatigue. The movements were corrected during the practise session, when necessary.

The video camera was set up on a tripod with angle adjustment allowing for variation in positioning for each test. Participants carried out the tests in a standardised order (Table 1), to ensure they were all assessed the same way. Standardised instructions were used for each test, for all participants (see Appendix 1 for example instructions of Test 1).

Scoring movement control impairments

Scoring using criteria detailed in Appendix 2 was completed for each participant from the video recording at a later date. Participants were judged against each criterion on a pass or fail basis. The videos were transferred onto a password protected computer where each participant was identified by a number and were edited using Microsoft Movie Maker (2014). The same rater (KR) viewed the videos of the participants' movement tests on two separate occasions, three weeks apart, in order to establish intra-rater reliability of rating performance. On both occasions, each participant's ability to control movement was assessed using the movement evaluation criteria (Appendix 2), which were posed as questions to the rater, to identify impairments. The rater was a novice to movement screening and received one training session, followed by practise sessions on colleagues.

Data Management

During the scoring process the investigator identified criteria that could not be evaluated due to being unable to see the appropriate part of the body clearly on the video. These criteria were, therefore, not included for analysis (please see the five boxes blacked out of the table in Appendix 2, and the footnote, for the excluded criteria); 23 of the 24 criteria were used, with 4 additional sub-components (left/right) removed (see Appendix 2).

Statistical Analysis

Data analysis was based on the three movement control tests making up the screening tool, with one test having been performed twice (to obtain video recordings from two different views), making four tests in total. Fourteen of the 23 movement control criteria were broken down into sub-components, parts a) and b) of a criterion (e.g. Test 1, side view, criterion 1) and left-side and right-side (e.g. Test 1, front view, criterion 1), making 37 sub-components in total. These sub-components can be seen clearly in the Table in Appendix 2.

Statistical analysis was carried out using IBM SPSS version 22.0, Microsoft Excel and Stata. Percentage agreement was measured and presented for overall test scores as a combination of test criteria and for each criterion as a whole (i.e. left and right legs combined/a) and b) combined). The intra-rater reliability of overall test scores was assessed using Cohen's Kappa with 95% confidence intervals. Kappa describes the "proportion of agreement corrected for chance" and is, therefore, believed to be more robust than percentage agreement (Cohen, 1960). The overall score for each participant is the sum of all criteria of each test. Intra-rater reliability is also presented for overall scores of each test. Cohen's kappa with 95% confidence intervals was also used to assess the intra-rater reliability of each individual criterion of each test. Where criteria of a test have subcomponents (e.g. left leg and right leg), Kappa is presented for both sides. When interpreting the strength of Kappa values greater than 0.80 are generally considered excellent, values between 0.60 and 0.80 substantial, values between 0.40 and 0.60 moderate and values below 0.40 poor to fair (Portney and Watkins, 2009).

Results

Percentage agreements for each test were: 96.3% for Test 1 (side view), 94.8% for Test 1 (front view), 97.3% for Test 2 and 93.0% for Test 3. Overall percentage test agreement, as a combination of all test questions, was 95%.

Kappa values indicated moderate and fair agreement of overall scores for Tests 1-side, 2 and 3 combined (κ=0.429), and 1-side, 1-front, 2 and 3 combined (κ=0.351), respectively. Substantial to excellent agreement was established for the scores of each individual test. Table 2 shows the Kappa values for intra-rater reliability for each test and scores for the tests combined.

Table 2: Intra-rater reliability of overall test scores: Cohen's kappa with 95% confidence intervals

Test	Kappa (95% CI)
1 – side view (1S)	0.635 (0.509-0.817)
1 – front view (1F)	0.731 (0.537-0.846)
2	0.900 (0.711-0.950)
3	0.600
1S, 2 and 3 combined	0.429 (0.250-0.514)
1S, 1F, 2 and 3 combined	0.351 (0.205-0.444)

Individual criterion analysis

More detailed analysis of individual criteria within each of the three tests was undertaken. Although it is useful to note the reliability of overall scores, this does not take into account agreement/disagreement within the tests. Analysis of each criterion reveals which criteria were most and least reliable.

The percentage agreement for criteria ranged from 83.9-100.0% (mean overall agreement 95.5%), indicating excellent agreement for all criteria (Table 3). Furthermore, only four of the 23 criteria had less than 90% agreement (Test 1, side view, criterion 6; Test 1, front, criterion 2; Test 3m criterion 4; Test 3, criterion 5), with nine criteria achieving 100% agreement (Figure 2 & Table 3). Agreement varied slightly between tests, with Test 3 being the least reliable; two criteria had agreement below 90% (4 and 5), and overall agreement was lowest at 93.0%. Percentage agreement for intra-rater reliability for each test criterion is illustrated clearly in

Figure 2, showing that most values close to or on the outer edge of the radar graph indicating very good agreement.

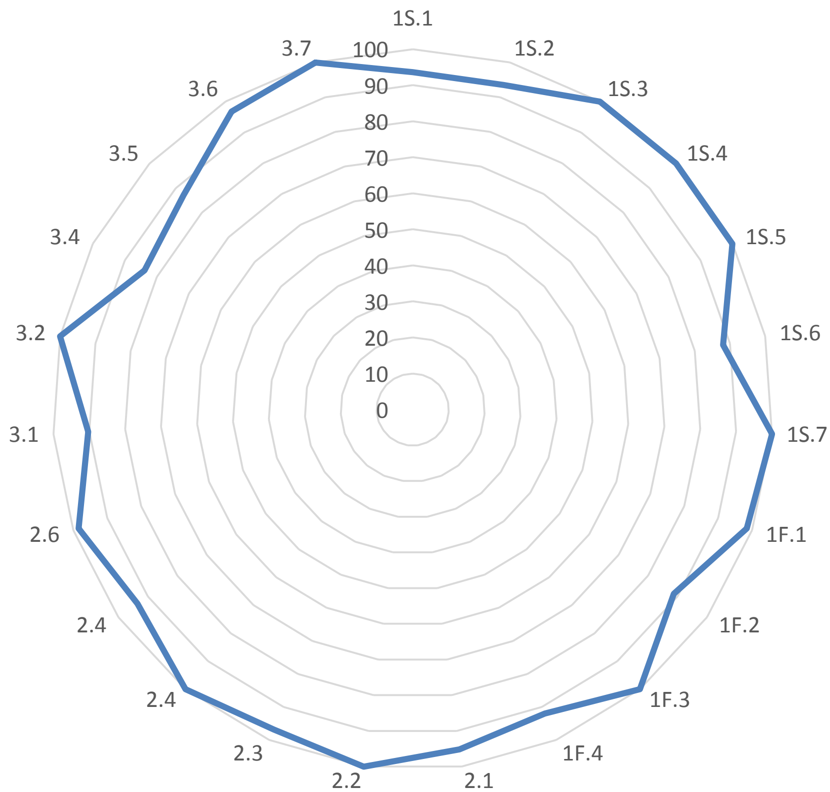

Figure 2: Radar graph showing percentage agreement (0-100 %) for individual test criteria. Starting at the top of the graph and going clockwise are the results for Test 1S (Test 1, side view) with 7 criteria, then 1F (Test 1, frontal view) with 4 criteria, Test 2 with 6 criteria and Test 3 with 7 criteria.

Cohen's kappa with 95% confidence intervals and percentage agreement for intra-rater reliability of individual criteria are presented in Table 3. Kappa values ranged from 0.00-1.00, with 17/37 criteria having κ>0.70 and three of the four tests containing values of 1.00. Eleven of the 37 criteria had κ<0.70, with two values of 0.00. A further 9/37 that were 100% agreement were κ=N/A (reason explained in the discussion). Some criteria had particularly good reliability (e.g. Test 2, criterion 6a, κ=0.902), whereas some criteria were less robust with lower reliability values (e.g. Test 3, criterion 5R, κ=0.431).

Table 3: Intra-rater reliability for criteria: Cohen's kappa with 95% confidence intervals and criterion percentage agreement (n=31 except *n=27)

Test	Criterion	Left-hand side/A** Kappa (95% CI)		Right-hand side/B** Kappa (95% CI)		Criterion agreement (%)
1 - side	1	0.693	(0.426-0.960)	N/A	N/A	93.6
	2			0.635	(0.180-1.000)	93.5
	3			N/A	N/A	100.0
	4	N/A	N/A	N/A	N/A	100.0
	5		N/A	N/A*		100.0
	6	0.724*	(0.432-1.000)	0.744	(0.517-0.970)	88.0
	7		1.000 (1.000-1.000)			100.0
1 - front	1	1.000	(1.000-1.000)	0.652	(0.022-1.000)	98.4
	2	0.597	(0.260-0.934)	0.737	(0.457-1.000)	88.7
	3	1.000	(1.000-1.000)	1.000	(1.000-1.000)	100.0
	4	0.728	(0.480-0.976)	0.933	(0.804-1.000)	92.0
2	1	0.832	(0.611-1.000)	0.000	(-1.000)	95.2
	2	1.000	(1.000-1.000)			100.0

	3	0.839	(0.534-1.000)			96.8
	4	1.000	(1.000-1.000)	1.000	(1.000-1.000)	100.0
	5	0.716 (0.350-1.000)				93.5
	6	0.903	(0.716-1.000)	1.000	(1.000-1.000)	98.4
3	1	0.773 (0.530-1.000)				90.3
	2	N/A	N/A	N/A	N/A	100.0
	4	0.676 (0.417-0.936)				83.9
	5	0.523	(0.114-0.932)	0.431	(-0.024-0.887)	87.1
	6	0.000	(-1.000)	0.652	(0.022-1.000)	96.8
	7	N/A	N/A	N/A	N/A	100.0

**See Appendix 2 for definition of A/B in each case

NB. Criteria 2 left, 3 left (test 1-side), 2 right, 3 right (test 2) and 3 (test 3) were not included in the analysis, for reasons explained in the methodology

Discussion

A movement control screen assessment tool comprising three simple tests has been described and has shown to be suitable for conducting in a community setting. Acceptable intra-rater reliability of observations repeated on two days from video recordings of the 'Movement control impairment screen for older people in community settings' was found. The strength of reliability was influenced by whether overall values for the screen, the three tests comprising it or individual test criteria were assessed and which statistical analysis was used (percentage agreement or Kappa).

An individual's overall score is generally used when assessing functional ability and movement control, e.g. for the Functional Movement Score (Teyhen et al., 2012; Gribble et al., 2013) or the Foundation Matrix (Mischiati et al, 2015). Percentage agreement was excellent for the present movement screen as a whole (95%) and overall values for each of the three tests ranged from 93-97%. Agreement for each of the test criteria revealed those that were least reliable, with four criteria showing values between 83% and 87% (Fig 2). The overall Kappa values for each of the three tests were fair to moderate (Table 2), whereas reliability of individual criteria was more variable, with the majority having substantial to excellent reliability (Table 3). It is evident that analysis of the screening tool as a whole, and of each test, does not allow weaknesses in specific criteria to be revealed, so analysing each individual criterion offers the opportunity to modify and refine the screening tool to make it more robust.

Clarity of Test Criteria

The least reliable test criteria may have been due to the way the criteria were expressed on the assessment form, so these were examined for clarity. Precise instruction of at which point during the test rater should make their judgement to assess the individual's movement would influence the accuracy and reliability of a test but this timing of the judgement was not always clear. For example, in Test 3 (chest rotation with neutral head and pelvis), criteria 1 and 4 assessed control of the head and pelvis whilst rotating the chest. It was noted that some participants were able to maintain a neutral head and/or pelvis at the beginning of the movement but lost control as the movement progressed. This caused difficulty when deciding whether the participant should pass or fail and although the reliability measurements were substantial (Kappa values 0.773 and 0.676, respectively), they were amongst the lowest of all test criteria (Table 3). It is proposed that an additional benchmark with a clear description of precisely when the rater should score the movement would improve reliability, e.g. the benchmark could be 30 degrees of movement, so the person would be asked to rotate to 30 degrees and the question would be “Can they keep the head facing forward and control/prevent the head from turning to follow the upper trunk rotation to 30 degrees?”. This would ensure that scoring is made at the same point, thus enabling more consistent comparison between assessments i.e. benchmark. Furthermore, some individuals were able to maintain a neutral head and/or pelvis when rotating the chest to one side and not the other but the scoring criteria only allowed pass or fail in general. It would be beneficial to assess turning to both sides in order to determine asymmetry. The specific questions would therefore be: “Can they keep the head facing forward and control/prevent the head from turning to the right to follow the upper trunk rotation at 30 degrees to the right?”, with the same question relating to the left side.

In Test 3, criteria 5 and 6 also showed relatively low reliability values compared with other criteria (range: κ=0.000-0.652). Criterion 5 assessed the ability to keep the knees stationary during chest rotation. The criterion did not state whether the rater should be assessing the movement of the right and left knees, or assessing knee control when turning to the right or left. It is likely that the investigator interpreted this criterion differently between observation sessions, producing low reliability. The description of the criterion needs to be more specific to avoid confusion, e.g. in the case of Criterion 5, instructing the rater to assess movement of one or both knees when turning to the right and then when turning to the left, giving separate boxes for rating each knee and during turning to each side. Since it is important to establish any differences in movement control between sides of the body and when turning to either side, it is suggested that both factors be accommodated by modifying the criteria in a revised screening tool, as indicated by the more specific questions above.

For Test 3, criterion 6 led to similar problems in relation to which side of the body was being observed. The criterion assessed the ability to maintain equal weight-bearing during chest rotation by asking: “Can they keep equal weight on both feet and prevent any lateral movement of the pelvis or shoulders during the chest rotation?” This did not specify whether assessment should be made for the left side and right side, or when turning to the left or right. Furthermore, the investigator had difficulty interpreting control of weight-bearing, perhaps due to some participants’ clothing, but the pelvic and should movements would be difficult to discern by eye without the use of additional benchmarks, such as joint markers, but these would bring additional complexity and the purpose is for the tests to be simple to perform and rate. It may therefore be best to exclude this criterion about weight bearing.

For Test 1 (Sit-to-stand plus arm lift), criteria could also be modified to improve reliability. For example, the criteria that assess the ability to prevent shoulder elevation and depression during arm lift and lowering could define ‘hitching’ and ‘dropping’ more clearly. It became apparent that some participants did not achieve a mid-position of the shoulders so it was sometimes difficult to assess their shoulder control. However, different individuals may have a different achievable mid-position, which is still appropriate for testing control, as long as they do not hitch or drop from their achievable mid-position at the start of the test. Furthermore, some participants moved their shoulders back to mid-position at the end of the movement (after dropping), upon hearing the instruction ‘lower your arm down slowly without letting your shoulder drop’. They had begun lowering their shoulder before hearing the full instruction, which would essentially mean they had failed the test but could have had the ability to pass it. Perhaps modifying the instruction to ‘without letting your shoulder drop, lower your arm down slowly’ may result in more accurate performance of the movement.

Participants' natural postures made it difficult to interpret some aspects of movement control and thus score certain criteria. For example, Test 2 (Trunk lean with knee bend and opposite arm lift) assessed the ability to prevent the head from dropping forward from neutral during the forward trunk lean. Some participants had naturally increased cervical flexion, which occasionally made this difficult for the investigator to interpret. Again, using each participant's achievable mid-position as a starting point (rather than anatomical neutral), and any flexion from that point meaning they would fail the test, may have been more appropriate. Some criteria relate to the ability to keep the knees apart and over the feet, which was difficult to assess in participants with valgus knees. Additional benchmarks taking individuals' natural postures into consideration could be incorporated into the criteria.

Comparison with other Movement Control Screening Reliability Studies

Existing studies have assessed the reliability of a number of movement screening tools, e.g. various studies have examined the reliability of the FMS, both between raters and within raters (Teyhen et al., 2012; Gribble et al., 2013). These studies predominantly report substantial intra-rater reliability, with some results comparable to those of the novice investigator in the present study. For example, Teyhen et al., (2012) reported substantial intra-rater reliability for composite scores for the FMS (ICC=0.74), whereas the individual components (tests) of the FMS demonstrated fair to moderate reliability (κ=0.29-0.74). Gribble et al., found excellent intra-rater reliability (ICC 0.946) for experienced athletic trainers, and substantial intra-rater reliability (ICC 0.771) for athletic trainers with little experience, indicating that intra-rater reliability strengthens with experience. These overall analyses in previous studies did not enable the reliability of the individual tests or test criteria to be assessed, or enable closer comparison with the present findings.

A reliability study (intra and inter-rater) of the Foundation Matrix, from which the present tests were developed involved experienced raters (Mischiati et al, 2015) so results are not directly comparable to the present study but their analyses also involved testing overall scores and individual criteria. Comparing the experienced raters form the Mischiati et al. (2015) study with the present novice rater, similar overall intra-rater percentage agreement was found: 97.5% for Rater 1 (14 years of experience of movement control testing), 93.9% for Rater 2 (7 years of experience) and 95% for the present rater. Kappa values were less comparable: 0.6-1.0 for Rater 1; -0.1-1.0 for Rater 2 and 0.35-0.90 for the present novice rater. The most experienced rater had the best results with both analyses. The content and number of variables in the two screening tools differed too much to enable comparison of results for the individual test components: the Foundation Matrix reliability study had 9 tests, with 40 criteria (Mischiati et al, 2015) whereas the present screening test for older people only had three tests with 23 criteria.

Direct comparison between studies is limited due to a number of differences in study designs, including experience of raters, test viewing methods and statistical analyses used. For example, although Teyhen et al., (2012) examined the intra-rater reliability of the FMS using novice (student) raters, as in the present study, the raters received 20 hours of FMS training. Furthermore, the tests were rated 48-72 hours apart, whereas the present study rated three weeks apart, perhaps a reason for the higher reliability of composite scores (ICC 0.74 compared to κ=0.351-0.429). However, all tests were observed in real-time by Teyhen et al., (2012), without the benefit of replaying video recordings.

Statistical Analysis using Kappa

Situations where Kappa does not always reflect agreement accurately between different sets of data have been demonstrated and its limitations can underestimate agreement (Feinstein & Cicchetti, 1990; Sim & Wright, 2005). The phenomenon of high agreement but low Kappa values was discussed by Feinstein and Cicchetti (1990), who highlighted that Kappa becomes inaccurate when the prevalence of one outcome is much greater than that of the other. For example, Kappa is undefined when both ratings consist entirely of one outcome, meaning there are no data relating to the random chance of the other outcome occurring. In the present study there were situations when the rater scored all participants "pass", on both occasions. Therefore, there were no data relating to "fail", so agreement could not be assessed across both outcomes. In such cases, Kappa was presented "N/A", despite 100% agreement (Table 3). Furthermore, there were some situations where kappa was zero, suggesting that agreement was no better than the agreement expected by chance. These situations arose when one rating included only a single outcome but the second included both. Therefore, the information on the chance probability of the second outcome was limited, leading to high chance probability and low Kappa, despite high agreement. An example from the present study is Test 2, criterion 1b: The rater scored "pass" for all participants on one occasion, however, the other occasion included one score of "fail". Therefore, there was total disagreement when the rater scored "fail". Further investigation is required to accurately establish agreement in such situations. Despite these limitations of using Kappa values, they are included in the present paper to enable comparison with other studies of movement screening tools that used this statistical test.

Limitations of the Present Study

Previous research has suggested that it is often difficult to achieve good reliability from visual material, although judgments may be enhanced with sufficient training (Luomajoki et al., 2007). It is acknowledged that the investigator in the present study had little experience and training on movement control testing, which included learning and teaching the exercises, and practice providing feedback to correct movement quality/technique. However, limited training can reflect clinical practice, where time and resources may not be available for extensive training. It must be considered that a more experienced rater is likely to demonstrate stronger intra-rater reliability, as in Gribble et al. (2013). Padua et al. (2014) reported 'seven steps for developing and implementing a preventive training program', aimed at injury prevention, and highlighted the importance of training assessors in order to attain high levels of competency.

The video camera was repositioned for each test and sometimes had to be adjusted between participants due to height differences. Errors occurred occasionally by not readjusting the camera between participants, so that a participant was not entirely in the frame. If a movement could not be seen a missing data point was entered, affecting the participant's overall test score.

The order of tests was a potential limitation of the study, as this was not randomised, so there could have been a learning effect across or fatigue could have occurred by the third test.

The present study included only one investigator to assess intra-rater reliability. In order for the results to be generalizable, more observers with different levels of experience need to be studied. The reliability of performance of the test manoeuvres by participants also needs to be examined, as well as inter-rater reliability and real-time observation, in addition to retrospective analysis of video recordings.

Despite limitations, intra-rater reliability was acceptable and could be further improved by removing or modifying test questions that could not be answered accurately, or at all.

Future Research

The present study has indicated specific criteria that could be removed or modified in order to refine the 'Movement control impairment screen for older people in community settings' and make it more robust. Future research could, therefore, begin with establishing the reliability of a refined screening tool. Furthermore, inter-rater reliability should be examined to establish whether the tool is reliable when utilised by different people. The reliability of observation and scoring real-time needs to be established, as the video reliability findings may not translate to the real-time situation. The tool is designed for use in the community and there may be some situations when real-time testing may be necessary. For research purposes, retrospective rating may be beneficial to allow more rapid testing when a group of participants may need to be tested at one location within a short time period. However, when testing in people's homes, there may be occasions when filming is not possible, e.g. if space is limited. Also, real time evaluation would allow observation of movement control from all angles, which may enhance the tool.

The validity of the screening tool also needs to be established by comparing observational scores with an objective gold standard test of movement, such as 3-dimensional motion analysis.

Finally, it would be beneficial to apply the screening tool to different populations, such as individuals in other age groups, less active individuals and males. This would indicate whether the screening tool can be applied across a range of people with varying functional abilities.

Conclusions

The ‘Movement control impairment screen for older people in community settings’ has demonstrated acceptable intra-rater reliability. The findings indicate utility of the screening tool in the assessment and identification of movement control impairments in active older people. Recommendations have been made for the refinement of some criteria in order to make the screening tool more robust and thus increase reliability.

Acknowledgements

The authors thank the participants from golf clubs in Hampshire for their valuable time and enthusiasm; those who helped with recruitment, Chris Cheesman (Corhampton GC), Sarah Moore (physiotherapist, Cams Hall GC), Moira Hoath (Hayling Island GC) and Radio Solent; Simon Brown for his support throughout the data collection process; Dr Sean Ewings and Dr Sam Connolly for statistical support; and Isabel Herrick and Charlotte Buckley, fellow MSc students to KR, for assistance with recruitment and screening in relation to their own projects.

Conflict of Interest

Disclosure: Mark Comerford and Sarah Mottram are Co-Directors of Movement Performance Solutions Ltd, which educates and trains sports, health and fitness professionals to better understand, prevent and manage musculoskeletal injury and pain that can impair movement and compromise performance in their patients, players and clients. None of the other authors has any conflict of interest to declare.

References

Berg, K., Wood-Dauphinée, S., Williams, J.I., & Gayton, D. (1989). Measuring balance in the elderly: Preliminary development of an instrument. *Physiotherapy Canada, 41*(6), 304-311.

Buford, T.W., Donovan, J.L., Marzetti, E., Wohlgemuth, S.E., Vandenborne, K., Pahor, M.,…Manini, T.M. (2012). Age-related differences in lower extremity tissue compartments and associations with physical function in older adults. *Experimental Gerontology, 47,* 38-44. doi: 10.1016/j.exger.2011.10.001

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46. doi: 10.1177/001316446002000104

Feinstein, A.R., & Cicchetti, D.V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology, 43*(6), 543-549.

Gribble, P.A., Brigle, J., Pietrosimone, B.G., Pfile, K.R., & Webster, K.A. (2013). Intrarater Reliability of the Functional Movement Screen. *Journal of Strength and Conditioning Research, 27*(4), 978-981. doi: 10.1519/JSC.0b013e31825c32a8

Heesch, K.C., van Uffelen, J.G.Z., van Gellecum, Y.R., & Brown, W.J. (2012). Dose-response relationships between physical activity, walking and health-related quality of life in mid-age and older women. *Journal of Epidemiology and Community Health, 66,* 670-677. doi: 10.1136/jech-2011-200850

Ikezoe, T., Mori, N., Nakamura, M., & Ichihashi, N. (2010). Age-related muscle atrophy in the lower extremities and daily physical activity in elderly women. *Archives of Gerontology and Geriatrics, 53,* e153-e157. doi: 10.1016/j.archger.2010.08.003

Lephart, S.M., Smoliga, J.M., Myers, J.B., Sell, T.C., & Tsai, Y. (2007). An eight-week golf-specific exercise program improves physical characteristics, swing mechanics, and golf performance in recreational golfers. *Journal of Strength and Conditioning Research, 21*(3), 860-869.

Loland, N. (2002). Reliability of the physical activity scale for the elderly (PASE). *European Journal of Sport Science, 2*(5), 1-12.

Luomajoki, H., Kool, J., Bruin, E.D., & Airaksinen, O. (2007). Reliability of movement control tests in the lumbar spine. *BMC Musculoskeletal Disorders, 8*, 90. doi: 10.1186/1471-2474-8-90

McNeil, W. (2014). The Double Knee Swing Test – a practical example of The Performance Matrix Movement Screen. *Journal of Bodywork and Movement Therapies, 18*(3), 477-481. doi: 10.1016/j.jbmt.2014.05.008

Mischiati CR, Comerford M, Gosford E, Swart J, Ewings S, Botha N, Stokes M, Mottram SL. (2015) Intra and Inter-Rater Reliability of Screening for Movement Impairments: Movement Control Tests from the Foundation Matrix. *Journal of Sports Science and Medicine* 14, 427-441

Monnier, A., Heuer, J., Norman, K. & Äng, B.O., (2012). Inter- and intra-observer reliability of clinical movement-control tests for marines. *BMC Musculoskeletal Disorders, 13*, 263. doi: 10.1186/1471-2474-13-263

Morley, J.E. (2012). Sarcopenia in the elderly. *Family Practice, 29*, i44-i48. doi: 10.1093/fampra/cmr063

Padua, D.A., Barnett, F., Donaldson, A., de la Motte, S., Cameron, K.L., Beutler, A.I., Marshall, S.W. (2014). Seven steps for developing and implementing a preventive training program.

Lessons learned from JUMP-ACL and beyond. *Clinics in Sports Medicine*, *33*(4), 615-632. doi: 10.1016/j.csm.2014.06.012

Podsialdo, D., & Richardson, S. (1991). The timed “up & go”: a test of basic functional mobility for frail elderly persons, *Journal of the American Geriatrics Society, 39*(2), 142-148.

Portney, L.G., & Watkins, M.P. (2012). *Foundations of Clinical Research: Applications to Practice* (3rd ed.). Prentice Hall: New Jersey.

Samuel, D., Rowe, P., Hood, V., & Nicol, A. (2011). The biomechanical functional demand placed on knee and hip muscles of older adults during stair ascent and descent. *Gait and Posture, 34,* 239-244. doi: 10.1016/j.gaitpost.2011.05.005

Siegenthaler, K.L., & O'Dell, I. (2003). Older golfers: Serious leisure and successful aging. *World Leisure Journal*, *45*(1), 45-52. doi: 10.1080/04419057.2003.9674304

Sim, J. & Wright, C.C. (2005) The kappa statistic in reliability studies: Use, interpretation and sample size requirements. *Physical Therapy, 85*(3), 257-268.

Teyhen, D.S., Shaffer, S.W., Lorenson, C.L., Halfpap, J.P., Donofry, D.F., Walker, M.J., Childs, J.D. (2012). The functional movement screen: A reliability study. *Journal of Orthopaedic & Sports Physical Therapy 42*(6), 530-537. doi: 10.2519/jospt.2012.3838

Torres-Ronda, L., Sanchez-Medina, L., & Gonzalez-Badillo, J.J. (2011). Muscle strength and golf Performance: A critical review. *Journal of Sports Science and Medicine*, *10*(1), 9-18.

Tinetti, M.E. (1986). Performance-oriented assessment of mobility problems in elderly patients. *Journal of the American Geriatrics Society*, *34*(2), 119-126.

Wang, C., & Bai, L. (2012). Sarcopenia in the elderly: Basic and clinical issues. *Geriatrics and Gerontology International*, *12*, 388-396. doi: 10.1111/j.1447-0594.2012.00851.x

Appendix 1: Instructions for Test 1

Test 1 – Sit-to-stand and arm lift

Starting position:
Sit on the chair with both feet on the floor hip width apart. The thighs should face forwards and the knees slightly wider than the feet. Sit up straight.

1. Put your hands on the front edge of the chair

2. Lean forwards at the hips keeping your back straight

3. Push up using your hands until standing upright keeping your knees over your feet as you stand up (keep the knees apart during the sitting to standing movement)

4. Put both shoulders in mid- position
   - shrug your shoulders up and drop them down – mid-position is half way in between

5. Don't let the shoulder blades move at all

6. With your thumb pointing up and palm facing in, lift your left arm up in front of you to horizontal, keeping your shoulder blade still in neutral
   (hold for 3 seconds)

7. Lower your arm down slowly without letting your shoulder blade drop

8. With your thumb pointing up and palm facing in, lift your right arm up in front of you to horizontal, keeping your shoulder blade still in neutral
   (hold for 3 seconds)

9. Lower your arm down slowly without letting your shoulder blade drop

Appendix 2: Movement Evaluation Criteria

Test 1 – Sit-to-stand + arm lift **Movement Control Evaluation** (SIDE VIEW)	**Score 1 if criteria failed**		
Upper Back			
1. Can they keep the upper back straight and the collar bones up (the chest should not drop into flexion) during: a) The forward lean to a standing position? b) The horizontal arm lift and lower? *Score 1 for impairment in the shaded box during either question a) or b)*			
Shoulder *Place score of 1 in shaded box for impairment on either left or right* *Indicate impairment on left and/or right (x) in clear boxes*	**Lt**	**Rt**	
2. Can they keep the shoulders level and steady (prevent the shoulder hitching into elevation) during the horizontal arm lift?			
3. Can they keep the shoulders level and steady (preventing the shoulder dropping down from the neutral position) during the arm lowering action?			
4. Can they keep the thumb pointing towards the ceiling to control/prevent the shoulder from turning into medial rotation as the arm is lifted to horizontal? (the palm should be facing inwards)			
Low Back / Pelvis			
5. Can they keep the low back straight (not rounding into flexion) during the forward trunk lean?			
6. Can they keep the low back straight (not arching into extension) during: a) The forward trunk lean? b) The movement from sitting to standing?			
Hip			
7. Can they fully straighten at the hips when they stand up? (There should be a vertical line from the trunk through the legs. The trunk should not be leaning forwards or the hips pushed backwards)			
Total out of 7 *(score of one for impairment on either Lt or Rt and a or b)*			
Test 1 – Movement Control Evaluation (FRONT VIEW)	**Lt**	**Rt**	
Shoulder			
1. Can they keep the shoulders level and steady (control / prevent the shoulder blade hitching into elevation) during the horizontal arm lift?			

2. Can they keep the shoulders level and steady (control / prevent the shoulder blade from dropping down from the neutral position) into downward rotation or depression during the arm lowering action?			
3. Can they keep the thumb pointing towards the ceiling to control/prevent the shoulder from turning into medial rotation as the arm lifts to horizontal? (the palm should be facing inwards)			
Hip			
4. Can they keep the knees apart (over the feet, preventing the knees from collapsing in towards each other) during the sit to stand movement?			
Total out of 4 *(score of one for Lt or Rt)*			

Test 2 – Trunk lean with knee bend and opposite arm lift **Movement Control Evaluation** (SIDE VIEW)	**Lt**	**Rt**	
Neck			
1. Can they control/prevent the head from dropping forward from neutral (straight line with the trunk) during: b) the trunk forward lean? c) arm lifting and lowering?			
Shoulder			
2. Can they keep the shoulder and trunk steady on the weight-bearing arm?			
3. Can they keep the non-weighting shoulder steady as the arm is lifted and reaches out?			
4. Can they keep the thumb pointing towards the ceiling as the arm lifts to horizontal? (the palm should be facing inwards)			
Low Back / Pelvis			
5. Can they keep the back straight and not let it round out into flexion during the forward lean?			
6. Can they keep the pelvis facing the front (not rotating) during: a) The foot lift? b) The opposite arm lift?			
Total out of 6 *(score of one for impairment on either Lt or Rt and a or b)*			

Test 3 – Chest rotation with neutral head and pelvis **Movement Control Evaluation** (FRONT VIEW)	**Lt**	**Rt**	
Neck			
1. Can they keep the head facing forward and control/prevent the head from turning or rotating to follow the upper trunk rotation?			
Shoulder			
2. Can they keep their shoulders level and in a neutral mid position (prevent hitching into elevation) during trunk rotation?			
Low Back / Pelvis			
3. Can they keep the low back straight and prevent arching into extension during the chest turn?			
4. Can they keep the pelvis facing the front and not allow it to rotate to follow the upper body (no rotation of the pelvis)?			
Hip			
5. Can they keep the knees apart and stationary during the chest turn? (no inward or outward movement of the knees across the line of the feet)			
6. Can they keep equal weight on both feet and prevent any lateral movement of the pelvis or shoulders during the chest rotation?			
Lower Leg / Foot			
7. Can they keep the feet facing forward with the inside edges of the feet parallel and prevent the foot from turning in or out during the chest turn?			
Total out of 6 *(score of one for Lt or Rt)*			

NB. Excluded criteria (five boxes blacked out) were: Test 1 (side view) criteria 2 left and 3 left; Test 2, criteria 2 right and 3 right; and Test 3, criterion 3. These criteria were removed for reasons explained in the data management section of the methodology.