The inter and intra-rater reliability of the Netball Movement Screening Tool

Rebecca Vanweerd

A thesis submitted to Auckland University of Technology in fulfilment of the requirements for the degree of Masters in Philosophy (MPhil)

2013

School of Rehabilitation and Occupational Studies

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ATTESTATION OF AUTHORSHIP

"I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for publication for the qualification of any other degree or diploma of a university or other institution of higher learning, except where due acknowledgment is made in the acknowledgements".

Signed: …………………………………………………. Date: ………………..

ACKNOWLEDGEMENTS

I would like to firstly thank and acknowledge my incredible family, who supported me un-conditionally throughout this whole process. Your love and support throughout all walks of life gives me the strength and motivation to always strive to give my best in everything I do. This thesis would not have been possible without you. An additional thank you goes to my sister Kate, who agreed to be the model for the photos throughout my thesis.

Secondly, I would like to thank my supervisors Duncan Reid and Peter Larmer, whose leadership, guidance, advice, patience and vast experience made writing this thesis infinitely easier. Being given the opportunity to work closely with people within your profession that you look up to and hold in high regard is amazing. I have truly enjoyed the opportunity and experience of working with both of you.

To my boss Ian Lyness and colleagues at Mairangi Bay Physiotherapy Clinic, thank you for putting up with my 'very flexible' schedule and for your understanding of my time and energy commitments to my thesis. Being able to work in such a supportive, understanding and above all fun work environment part time over the past two and a half years has only done good things for me and my thesis!

A huge thank you to my other boss and great friend Brianne Law, who kept me on the road and on the payroll for Alpine Canada after I moved back to New Zealand to undertake the writing of my thesis. I can't thank you enough for your support of me and my on-going education and for your support and company as I studied in hotel rooms, hotel lobbies and hotel bars all over the world!

I would like to thank AUT for granting me the Faculty of Health and Environmental Sciences Postgraduate Fee Scholarship in 2012 that went a long way in supporting me and my student lifestyle of the marginally employed.

I would also like to thank the NZMPA for awarding me a grant through the NZMPA Education fund to assist me with course and study related costs. This financial help was truly beneficial.

A big thank you also to Greig Bramwell from Asics whose kind donation of shoes was much appreciated and very well received by the three lucky recipients.

Another big thank you to Drew South at AUT Akoranga's Library. Making sense of EndNote and finally learning how to search effectively made a huge difference to me.

Finally I would like to thank the schools; Westlake Girls, Pinehurst and Carmel College that took part in the study. A huge thank you to Gloria Kelly, Marcia Hardcastle and Tasha Sofaea, my direct contacts at each of the schools that made my recruiting of subjects possible, and to the coaches who allowed me to interrupt their training somewhat to screen the girls. Finally I would like to thank all the subjects and their parents who agreed to participate in the study.

ETHICAL APPROVAL

The University of Technology Ethics Committee (AUTEC) granted ethical approval for this research on 17/11/2011. Approval Number: 11/296 (see Appendix A).

ABSTRACT

Objective

To establish the inter and intra rater reliability of the Netball Movement Screening Tool (NMST), for screening in a cohort of adolescent female netball players.

Background

Netball is one of the most popular team sports in the Commonwealth. It is a fast-paced dynamic sport that can result in injury, especially to the knee and ankle. The most important factor in reducing costs and disability from sports related injuries is to prevent injuries from happening in the first place. To date there is no screening tool, specific to netball, which is used to identify faulty movement patterns that may be able to predict injury risk. The NMST has been developed to identify netball players with poor movement patterns. For a screening tool to be shown as a valid tool in identifying faulty movement patterns it must first be shown to be a reliable tool. This study investigates the inter and intra rater-reliability of the NMST in a group of secondary school netballers.

Methods

Forty secondary school netball players were recruited to take part in the study. Participants were screened with the NMST which consisted of; the Movement Competency Screen (MCS), Jump components, Star Excursion Balance Test (SEBT) and Active Straight Leg Raise (ASLR). Twenty subjects were screened simultaneously and independently by two raters to ascertain inter-rater agreement. Twenty subjects were scored by rater one on two occasions separated by a week to ascertain intra-rater agreement. Rater one was a physiotherapist with over 10 years of experience in physiotherapy and rater two was a final year physiotherapy student. Inter and intra-rater agreement was assessed for; overall NMST, MCS, Jump component, ASLR and SEBT overall scores and for the individual tests of the MCS and Jump components, utilizing the two-way mixed intra-class coefficient (ICC), standard error of measurement (SEM), minimal detectable change score (MDC) and weighted kappa statistics.

Results

No significant differences were found between the demographics of the inter and intra rater groups of subjects. ICC demonstrated excellent inter (ICC = 0.84; SEM = 0.25; MDC = 13.8) intra-rater (ICC = 0.96; SEM = 0.13; MDC = 10.5) reliability for overall NMST score and substantial – excellent (ICC = 1.0- 0.65) inter-rater and substantial-excellent intra-rater (ICC = 0.96- 0.79) reliability for the category scores of the NMST. Kappa statistic showed substantial-poor inter (k = 0.75-0.32) and intra-rater (k = 0.77-0.27) agreement for individual tests of the NMST.

Conclusion

The NMST was shown to have excellent inter and intra-rater agreement for overall score in a cohort of adolescent female netball players. Excellent-substantial inter and intra-rater agreement for the category scores of the NMST were shown, whereas individual tests of the NMST showed more variable agreement. Decreased variability of scoring and improvements in the MDC score are required to ascertain meaningful clinical change. Future studies need to be undertaken to establish if there is a link between lower NMST score and increased injury risk.

Chapter 1

Introduction

1.1 Pre-season Movement Screening

The most important factor in reducing costs and disability from sports related injuries is to prevent injuries from happening in the first place (Ostenberg & Roos, 2000; Trojian & McKeag, 2006). Identifying athletes who are deemed to be at an increased risk of sustaining injury, through the implementation of pre-season or pre-participation physical exam, has been a standard prerequisite in most athletic programmes for the past 30 to 40 years (Kibler, Chandler, Uhl, & Maddux, 1989; Samples, 1986; Saunders & Eggart, 1979). Originally these pre-participation screens were driven by medical doctors with the aim of obtaining information that may reduce the chances of severe injury and death (Kibler et al., 1989). Although the main aim of pre-participation screening remains the same, the content and structure of these screens is ever-changing in line with current research and thinking (Miller & Callister, 2009).

Over the years different risk factors have been proposed as injury inducing and screens have attempted to identify athletes harbouring the isolated risk factors such as; flexibility, strength, range of motion and balance often around a single joint or aimed at a single pathology (McGuine, Greene, Best, & Leverson, 2000; McHugh, Tyler, Tetro, Mullaney, & Nicholas, 2006; Trojian & McKeag, 2006; Wang, Chen, Shiang, Jan, & Lin, 2006). Screens focusing on isolated risk factors often around a single joint have yielded inconsistent results especially in regard to their relationship with global injury (Butler, Plisky, Southers, Scoma, & Kiesel, 2010). It has been suggested that isolated testing that targets a single joint, muscle or an isolated risk factor, fails to link how the kinetic chain responds to this deficiency when it performs fundamental movement patterns that are used in; activities of daily living, sports and sport specific movements (Kritz, Cronin, & Hume, Forthcoming: 2014). Plisky, Rauh, Kaminski and Underwood (2006) proposed that tests which assess multiple domains of function (balance, strength, range of motion) simultaneously may improve the accuracy of identifying athletes at risk of injury. This evolution in thinking and the subsequent development of screens that assess multiple domains of function has now brought us into the realm of movement screening.

The recent shift in athletic screening focuses on the fundamental movement patterns of the individual athlete. This approach is underpinned by the assumption that deficits in functional movement patterns may increase susceptibility to injury and may also hinder athletic performance (Cook, Burton, & Hoogenboom, 2006; Kiesel, Plisky, & Voight, 2007; Minick et al., 2010; Schneiders, Davidsson, Horman, & Sullivan, 2011). Nagano et al. (2010) looked at the relationship between clinical physical measurements and knee motion during landing. Lower limb movement during landing, especially valgus, was found to be almost independent of clinical physical measurements and therefore necessary to evaluate on its own. These findings give weight to the use of screens that assess movement and function rather than relying on clinical measures such as range of motion and static foot alignment which were also assessed in the Nagano et al. (2010) study. Evaluation of isolated risk factors does not take into account how an athlete performs the functional multi-faceted movements required for sports (Kiesel et al., 2007). It has also been suggested that a fundamental change in motor control may occur after injury and that movement oriented tests may be better equipped to detect these changes (Kiesel, Plisky, & Butler, 2011). It has been therefore proposed that by simultaneously assessing multiple domains of function, athletes that are at risk of injury may be more accurately identified.

Several screens that assess multiple domains of function are being utilised clinically and in a research setting (Hewett et al., 2005; Kiesel et al., 2007; Plisky et al., 2006). Although all of these screens are ultimately looking to ascertain increased injury risk, each movement screen has its own individual focus. It may be that different screens or elements thereof, are better able to detect increased injury risks in different populations or sports. Having sport specific measurement tools which assess the domains of function pertinent to a particular sport may help to identify athletes who have an increased injury risk and allow for resources to be directed at those who will most benefit from them within that individual sport.

1.2 Netball

Netball is a fast paced dynamic team game involving; running, jumping, passing, catching, explosive power, pivoting, sudden changes in direction and accuracy. It is one of the most common sports in the Commonwealth (Steele, 1990) and has been shown by Sport and Recreation New Zealand (SPARC) to be the most played by women in New Zealand (SPARC, 2001). Netball New Zealand report over 138,000 registered participants in 2013 (Netball New Zealand, 2013) . Netball is the most popular team sport in Australia with the highest number of registered participants in any one sport (Hopper, Elliot, & Lalor, 1995) and an estimated one in seven females participate on a regular basis (McManus, Stevenson, & Finch, 2006). Although played by both genders, netball is a female dominated sport. The 2007/8 Active New Zealand Survey, showed that 88.3% of netball participants were female (SPARC, 2009). In New Zealand netball is one of only 9 targeted National Sporting Organisations (NSOs) which is given funding by the High Performance Sport New Zealand.

1.3 Netball Injury Incidence

Injury incidence and injury profiling studies specifically for netball are limited. Injury incidence for non-elite netball players has been reported at 5.4% (Hopper et al., 1995) and 14 injuries per 1000 playing hours (McManus et al., 2006), with injury incidence increasing with the level of competition.

Data from the Accident Compensation Corporation (ACC) of New Zealand shows a growth in the number of accepted claims from 2006-2011 for injuries caused by netball. In the year 2006, ACC accepted 18, 682 claims for netball injuries (ACC, personal communication, February 28, 2012). This number has continued to rise each year with 26,321 claims for netball accepted in 2012 (Johnston, 2013). The number of claims accepted by ACC continued to rise between 2009 and 2010 when player number actually decreased (Johnston, 2013). The highest number of injuries across the 6 years fell in the 10-14 year old age group followed by the 15-19 and the 20-24 year old age groups respectively (Johnston, 2013).

1.4 Netball Injury Profile

Studies investigating netball injuries show these mostly occur in the lower leg, especially in the knee and ankle (Flood & Harrison, 2009; Hopper & Elliot, 1993; Hopper et al., 1995; Otago & Peake, 2007). Ankle injuries have been shown to account for 19.3% (Flood & Harrison, 2009), 22% (Hopper & Elliot, 1993) and 84% (Hopper et al., 1995) of total injuries sustained. Whilst knee injuries have been shown to account for 37.4% (Flood & Harrison, 2009), 58% (Hopper & Elliot, 1993) and 8.3% (Hopper et al., 1995) of total injuries sustained with Anterior Cruciate Ligament (ACL) rupture alone accounting for 17.2% (Flood & Harrison, 2009) of total injuries sustained. Otago and Peake (2007) reported that 85.3% occurred in the lower limb, which accounted for 85.4% of total insurance costs.

1.5 Significance of the problem

With the large percentage of New Zealand females playing netball and the increased physicality of the sport (Johnston, 2013) the potential for injury and the long term sequel that can follow is high. Keeping this population healthy and in physical activity for life could prevent long term health problems such as; osteoarthritis, osteoporosis, obesity, heart disease and type II diabetes and the subsequent burden they place on the health care system (Brukner & Khan, 2012; WHO Consultation on Obesity, 1999). As previous injury has been seen as a deterrent for participation in exercise (Finch, 1996; Sale, 1991), preventing injuries from occurring in the first place will have many short and long term benefits for the athlete specifically, as well as for the healthcare system in general.

Flood and Harrison (2009) reported that 17.2% of hospital admissions for female netball players were due to ACL rupture. ACL rupture is a serious injury that often requires prolonged recovery, surgery and post-operative care and rehabilitation (Ford, Myer, & Hewett, 2003; Hewett et al., 2005). It is noted that 50% of post ACL-reconstructive athletes develop early osteoarthritis within the first 10 years post reconstruction, thus creating long term health issues and significant costs (Myklebust & Bahr, 2005; Myklebust et al., 2003). Therefore the prevention of ACL injury in netball is of utmost importance. The risk of ACL rupture is 4-6 times greater in female athletes when compared to their male counterparts whilst participating in the same sports (Hewett, Myer, Ford, & Slauterbeck, 2006). Although there have been many theories as to why such a large gender disparity exists, a consensus has yet to be reached and it appears to be somewhat multi-factorial in nature. Gender disparity has been described under three main categories; anatomical, hormonal and neuromuscular (Hewett et al., 2005). Of these three categories the only one that has been seen to be modifiable is neuromuscular patterning. Neuromuscular re-training to prevent faulty movement patterns has been touted as an intrinsic factor that may be able to be altered and reduce the incidence of ACL injury (Hewett et al., 2005) and may therefore play a part in injury prevention programmes for netball.

Adolescence appears to be the critical time for females to increase bone density (2009; Brukner & Khan, 2012). High intensity weight bearing exercise is important for promoting bone mass to aid in the prevention of osteoporosis in later life (2009; Brukner & Khan, 2012). It has been documented that adolescent females are less active compared to their male counterparts and children (2009; Australian Bureau of Statistics, 2009). Pain and injury are among the factors that contribute to decreased sport participation (Australian Bureau of Statistics, 2009). Injuries that prevent adolescents from participating in high intensity weight bearing exercise may have ramifications in the long term, leading to increased financial demand on the healthcare system. Therefore, preventing injury and encouraging ongoing participation in high intensity weight bearing exercise, such as netball, is critical for adolescent females.

To reduce long term health costs and disabilities from netball, preventing injuries from occurring is paramount. Being able to screen netball players for the movement patterns that may increase the risk of sustaining an injury would be beneficial. For a screen to be shown as a valid tool in identifying increased injury risk, it must first be shown to be reliable.

This thesis attempts to identify a reliable movement based screening tool for netball players. Following the Introduction section of this thesis, Chapter Two consists of a literature review that was undertaken to ascertain whether any reliable tools that could be used to predict injury risk in netball players already exists. Chapter Three outlines the methodology utilised to ascertain the inter and intra-rater reliability of the Netball Movement Screening Tool. Results of the reliability study are given in Chapter Four and these results are discussed in Chapter Five.

Chapter 2

Literature Review

This chapter provides findings from a systematic review that was undertaken to identify studies looking at the inter and/or intra rater reliability of movement screening tools, specific and non-specific to netball, aimed at identifying movement patterns that may predict injury risk. The identified studies were then rated for methodological quality rating utilising the modified Down's and Black (Downs & Black, 1998) for reliability studies. The methodological quality and results of the studies were then examined and compared across studies.

2.1 Methodology

The purpose of this literature review was to identify studies looking at the reliability of movement screening tools (specific to netball) used to identify movement patterns that may predict injury risk. A preliminary review was undertaken looking at this concept. No studies were identified that looked specifically at netball. As a result, a systematic review was undertaken to identify studies looking at the reliability of movement screening tools, non-specific to netball, used to identify movement patterns that may predict injury, in order to help develop a tool specific to netball. A thorough search was undertaken using the electronic databases subscribed to by Auckland University of Technology. These included; CINAHL Plus with Full Text, Health Source: Nursing Academic Edition, MEDLINE, SPORTDiscus with Full Text and Biomedical Reference Collection: Basic. Searches were performed between 28 July and 18 August 2012. An experienced health science librarian assisted with the development of the search. Full text studies that were written in, or translated into English and carried out on human participants were accepted. No limit was placed on publication date. The keywords used in isolation or combinations were; "pre-season screen*" OR "netball*", "pre-participation screen*", "move*base* screen*" , "move* screen*", "musculoskel* screen*","physical screen*", reliabil* and NOT cardi*. In addition to these keywords known movement based screening tools that have an intent to prevent injuries were also used namely; "functional movement screen", "star excursion balance test", "landing error scor* system", "knee valgus*" and "drop vertical jump". A manual search was also conducted on the reference lists of the relevant articles.

Studies were selected based on the following inclusion criteria; reliability studies assessing inter and/or intra-rater reliability of movement based screening tools aimed at establishing increased injury risk, healthy participants with no known musculoskeletal pathology and written in or translated into English. Studies looking at static tests alone, studies that focused on a single movement or single joint, studies that did not assess dynamic movement, screening tools that did not aim to ascertain increased injury risk and studies not written in or translated into English were excluded.

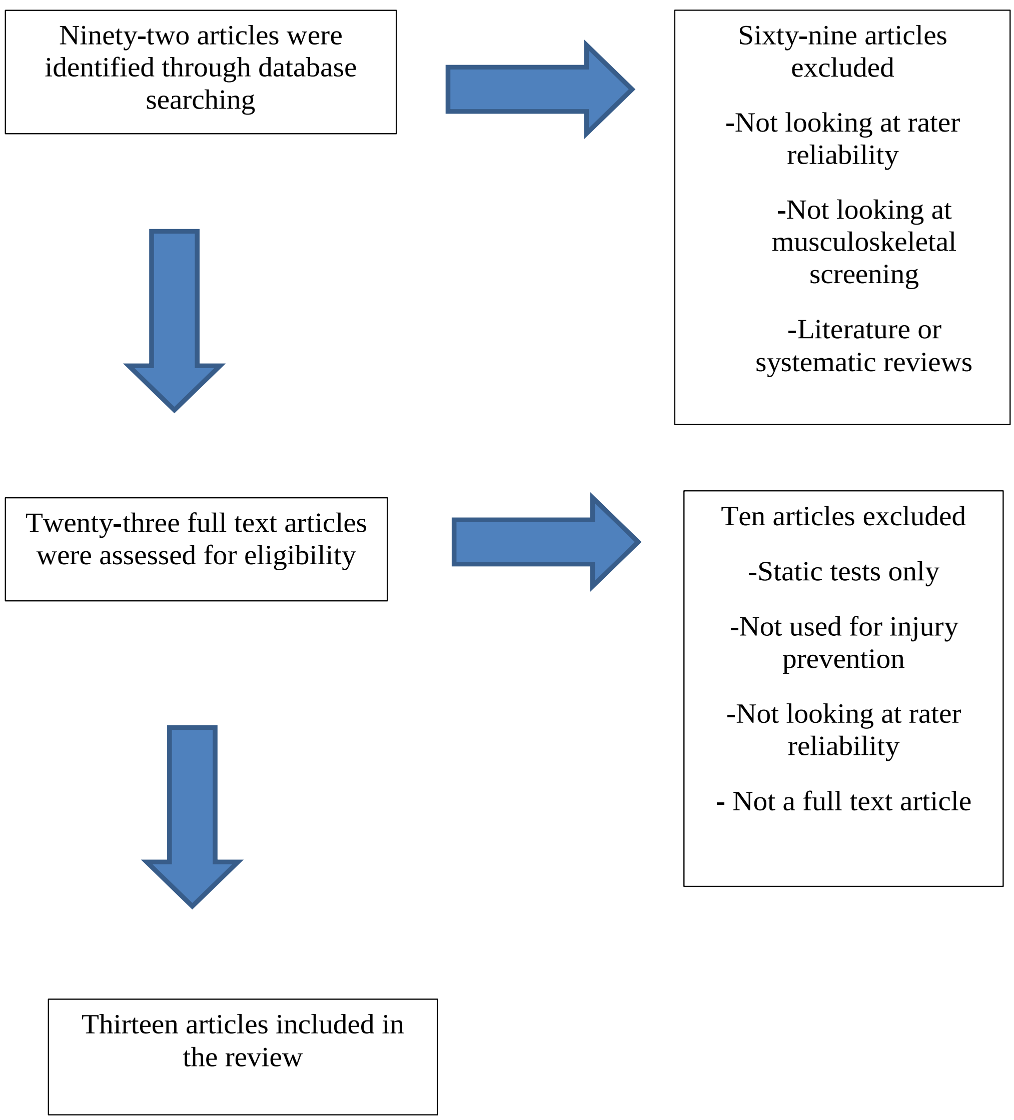

Figure 1: Flowchart of database search strategy

Ninety two studies were identified through database searching. Twenty three full text articles were assessed for eligibility. Ten of these articles were not deemed to have met inclusion criteria (see Appendix B). Thirteen articles were included in the review. One study (Karim, Millet, Olson, & Morganthaler, 2011) included a component of sports specific movement based screening as well as a battery of static tests. The movement based component of the screen only was utilised in this review. No articles were found looking at movement screening tools and netball players specifically.

2.2 Methodological quality rating

A modified Downs and Black for reliability studies (mD&B) (Appendix C) was used to ascertain methodological quality of the identified studies by the primary researcher. The Down's and Black Checklist for measuring study quality was developed to assess the methodological quality of both randomised and non-randomised studies of health care intervention (Downs & Black, 1998). As reliability studies have key fundamental methodological differences to both randomised and non-randomised studies involving a health care intervention, this checklist was modified to better assess the quality of studies looking at inter and/or intra-rater reliability. Table 1 shows how methodological quality was scored.

Table 1: Methodological quality rating scores for mD&B

	Excellent	Good	Fair	Poor
Inter or intra rater alone	≥ 16/18	14-15/18	11-12/18	≤ 10/18
Both inter and intra rater	≥ 17/19	15-16/19	12-14/19	≤ 11/19

Thirteen articles met inclusion criteria and were included in the review. Table 2 shows the mD&B score for the reviewed studies. An overview comparing the reviewed studies is seen in Table 3.

Table 2: Modified Down's and Black (mD&B) scores for reviewed studies.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	mD&B score
Ekegren et al, 2009	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	UTD	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	17/19
Schneiders et al, 2011	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	UTD	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	16/18
Karim et al, 2011	Yes	Yes	Yes	Yes	Yes	Yes	n/a	Yes	UTD	Yes	Yes	Yes	Yes	UTD	Yes	Yes	Yes	No	15/18
Onate et al, 2010	Yes	Yes	Yes	Yes	Yes	Yes	n/a	Yes	UTD	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	15/18
Onate et al, 2012	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	UTD	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	16/19
Padua et al, 2011	Yes	Yes	Yes	Yes	Yes	Yes	n/a	Yes	UTD	Yes	Yes	UTD	Yes	Yes	Yes	Yes	Yes	No	15/18
Minick et al, 2010	Yes	Yes	Yes	Yes	Yes	Yes	n/a	Yes	UTD	Yes	No	Yes	Yes	UTD	Yes	Yes	Yes	No	14/18
Padua et al, 2009	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	UTD	UTD	UTD	Yes	Yes	Yes	Yes	Yes	Yes	No	15/19
Teyhen et al, 2012	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	UTD	Yes	Yes	Yes	Yes	UTD	No	Yes	Yes	No	15/19
Kritz et al, (Forthcoming, 2014)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	UTD	Yes	UTD	UTD	Yes	Yes	Yes	Yes	Yes	No	14/19

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	mD&B score
Plisky et al, 2009	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	UTD	UTD	UTD	Yes	Yes	Yes	Yes	Yes	Yes	No	14/19
Hertel et al, 2000	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	UTD	UTD	UTD	UTD	Yes	Yes	Yes	Yes	No	No	13/19
Demura et al, 2010	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	UTD	UTD	UTD	Yes	UTD	UTD	Yes	No	No	12/19

For each question studies are scored; Yes (met criteria), Unable to determine (UTD) (not able to determine whether criteria was met) or No (criteria not met). A score of 1 is given when a question is answered with ‘Yes’ (aside from Q4 for which a yes is awarded a score of 2 as per the original Downs and Black checklist) and a score of 0 is given if the question is answered with ‘No’ or ‘Unable to Determine’. Studies that investigated both inter and intra-rater reliability were given a score out of 19 and studies that looked at only inter-rater or only intra-rater reliability were given a score out of 18. Studies are listed from highest to lowest overall methodological score and in alphabetical order for studies with the same methodological score.

2.21 Results of Modified Down's and Black

Two of the studies (Ekegren, Miller, Celebrini, Eng, & MacIntyre, 2009; Schneiders et al., 2011) were judged by the mD&B to have excellent methodological quality. Seven of the studies (Karim et al., 2011; Minick et al., 2010; Onate, Cortes, Welch, & van Lunen, 2010; Onate et al., 2012; Padua et al., 2011; Padua et al., 2009; Teyhen et al., 2012) were judged to have good methodological quality and four (Demura & Yamada, 2010; Hertel, Miller, & Denegar, 2000; Kritz et al., Forthcoming: 2014; Plisky et al., 2009) were judged to have fair methodological quality. The key areas that were consistently marked low in the mD&B for the reviewed studies were; Question 9, 10, 11, 12, 14 and 18. In many cases the information needed to answer these questions was not described in the methodology of the reviewed studies meaning that the question was marked as unable to determine and a score of zero was given.

Only one study (Demura & Yamada, 2010) was deemed to have selected subjects that were representative of the entire population from which they were recruited. These authors compared the demographics of their subjects to the average for Japanese males. This question (Question 9) was marked as 'no' and scored zero if the authors of the study did not identify their target population and the demographics that would allow comparison between the subjects and target population to be made. Most studies used volunteers or a convenience sample where this was not possible. Convenience sampling is a popular method of sampling as it is a fast inexpensive method of sampling and usually ensures subjects are within close proximity to the researcher (Urdan, 2010). However this form of sampling may not ensure subjects are representative of the entire population from which they were recruited. For the studies that utilised military personnel, it may be assumed that, if their target population is military personnel, a convenience sample of military personnel may be representative of all military personnel. However demographical data is required to support this assumption. Recruiting subjects that are representative of the target population is a challenge within reliability studies and needs to be addressed. It is possible that the reviewed studies did select subjects that were representative of the target population and failed to make this clear in their methodology.

Question ten asked whether assessors were adequately trained in standardised testing procedures and equipment. Of the five studies (Demura & Yamada, 2010; Hertel et al., 2000; Onate et al., 2012; Padua et al., 2009; Plisky et al., 2009) that scored a zero for this question , only one study (Onate et al., 2012) mentioned rater training. Onate et al (2012) used a novice rater that had read the FMS manual only once.

Question eleven ascertained whether or not assessors had similar backgrounds in the implementation and scoring of the screening tool of interest. This was deemed unable to determine in five studies (Demura & Yamada, 2010; Hertel et al., 2000; Kritz et al., Forthcoming: 2014; Padua et al., 2009; Plisky et al., 2009). Three studies (Minick et al., 2010; Onate et al., 2010; Onate et al., 2012) used raters with differing levels of experience and were marked zero for doing so.

Studies that scored a zero for question fourteen (Demura & Yamada, 2010; Karim et al., 2011; Minick et al., 2010; Teyhen et al., 2012) looking at familiarising subjects with the screening tool of interest and giving detailed instructions on how to perform the screening tool components, were all deemed to be unable to determine. This is an important question as a learning effect, which needs to be accounted for, may be associated with the screening tool. Performance of items of a screening tool may be sub-optimal if the subject is not given correct information or instructions, or does not understand the information or instructions that they are given.

Only one study (Ekegren et al., 2009) justified their sample size. Justification of sample size is important to determine whether results are statistically valid. This needs to be included in future studies.

The Kritz et al. (Forthcoming: 2014)study had methodology that was distinctly different to the other reviewed studies, as multiple raters and a very small number of subjects were used, rather than multiple subjects and a small number of raters.

The mD&B penalises studies that have differing training and experience, which may be a flaw in the tool, rather than a methodological flaw in the studies that it assesses the quality of. In fact a tool or tests that can be administered by raters with varying levels of training and experience is advantageous in a clinical setting. Some of the reviewed studies may have been penalised for this.

Table 3: Comparison of all reviewed studies

Study	Score	Screening Tool	Type of Study	Subjects	Raters	Video/Real Time	Analysis	Results
Ekegren et al, 2009	17/19 Excellent	Dynamic knee valgus	Inter-rater and Intra-rater	-40 adolescent females -Convenience sample from local soccer teams - Age:15 ± 1 years old -Height: 165 ± 6cm -Weight: 60 ± 8.5kg	-3 volunteers from a pool of 15experienced physiotherapists with similar clinical background and experience measuring dynamic knee valgus -Raters watched and scored independently	**Video Review** -Single viewing only, no pausing -15 seconds between subjects to record ratings	**Inter-rater** -Multi-rater kappa coefficient for **Intra-rater** -Percentage of agreement and kappa coefficient	**Inter-rater** -Substantial to excellent agreement at both time points; k = 0.80 and 0.77 **Intra-rater** -All 3 raters achieved substantial to excellent agreement; k = 0.80, 0.85 and 0.75
Schneiders et al, 2011	16/18 Excellent	FMS	Inter-rater	-A subset (28%) representative of the 209 healthy adults in main study -Convenience sample, targeted and recruited from tertiary	-2 raters -Same clinical experience -Same experience in implementation of FMS -Raters scored simultaneously	**Real time** -Three trials of each test	-ICC (3,1) for total FMS score -Unweighted kappa for individual items	-Total score: ICC = 0.97 -Individual items; 6/7 final scores showed excellent agreement; k = 0.86-1.00. 6/10 left/right sided

Study	Score	Screening Tool	Type of Study	Subjects	Raters	Video/Real Time	Analysis	Results
Schneiders et al, 2011				student population -Demographics only from total population (n=209) -Age: 21.9 ± 3.7 (18-40) years old -Height: 172.3 ± 9cms Weight: 72.8 ± 12kgs	and independently			scores showed excellent agreement; k = 0.84-1.0. All remaining showed substantial agreement; k = 0.73-0.80.
Karim et al, 2011	14/18 Good	Dynamic Posture, based on pilot 2006 Dance USA Annual Post-Hire Health Screen for Professional Dancers	Inter-rater	-30 female professional dancers - Convenience sample from 6 dance companies -Mean age: 24 (18-32) years old -Mean height: 164.08cms -Mean weight: 60.51 kgs	-4 raters -1 licenced physical therapist and 3 physical therapy students -Scored independently and simultaneously	**Real time** -Rater walked around dancer as they performed the movements -No information given on number of repetitions	-Percentage of agreement	-Percentage of agreement was classed as high (70-100%) for 7/10 tests -Other three tests had percentage of agreement between 40-70% -No explanation of classification given

Study	Score	Screening Tool	Type of Study	Subjects	Raters	Video/Real Time	Analysis	Results
Onate et al, 2010	15/18 Good	LESS	Inter-rater	-Nineteen female soccer players -Sample of convenience -Age: 19.58 ± .84 years -Height: 167 ± 5cms -Weight: 63.66 ± 10.11 kgs	-2 raters, certified athletic trainers -Expert rater involved in LESS development -Novice rater with one hour training. No previous implementation of LESS - Raters scored independently	**Video Review** -Viewed on personal laptops as many times as desired	-Kappa statistic for agreement on LESS items -ICC for overall LESS score	-Individual items moderate to excellent; k = 0.46-0.88. -4/15 perfect agreement -Overall score excellent reliability ICC = 0.84
Onate et al, 2012	16/19 Good	FMS	Inter-rater and Intra-rater	-19 physically active adults (12 males, 7 females) -Sample of convenience from local university and surrounding area - Age: 25.08 ± 3.12 (males)	-2 raters -Expert rater, 4 years' experience as certified athletic trainer and strength and conditioning specialist. FMS certified specialist	**Real time** -Raters attempted to stand with similar view of the subject -Subjects performed 3 repetitions of each movement	**Inter-rater** -ICC and SEM or total score and kappa and SEM for individual tests **Intra-rater** -ICC and SEM for total score and kappa and SEM for	**Inter-rater** -Total score: ICC = 0.98; SEM = 0.25 -Individual items k = 0.33-1.00 **Intra-rater** -Total score ICC = 0.92 -Individual items k = 0.16-0.84

Study	Score	Screening Tool	Type of Study	Subjects	Raters	Video/Real Time	Analysis	Results
Onate et al, 2012				25.29 ± 2.81 (females) years -Height: 181.41 ± 4.47 (males) 164.19 ± 5 (females) cms -Weight: 85.13 ± 7.72 (males) 60.59 ± 6.16 (females) kgs	-Novice rater, 3 years' experience as strength and conditioning specialist. Never scored FMS before. Read manual one time. -Raters scored simultaneously and independently		individual tests	
Padua et al, 2011	15/18 Good	LESS	Inter-rater	-43 healthy adult volunteers (24 females, 19 males) -Freshman at the US Military Academy - Height: 172.11 ± 6.85cm -Weight: 70.33 ± 10.11kg	-Three raters -Certified athletic trainers with 5+ years' experience -Same training in implementing and scoring LESS -Scored simultaneously and independently	**Real-time** -Subjects performed 4 trials -Specific items scored on specific trial	-ICC 2,1, SEM and 95% CI for LESS final score	-Rater 1 and 2; ICC = 0.81; SEM = 0.69 -Rater 1 and 3; ICC = 0.72; SEM = 0.79 -Combined; ICC = 0.79; SEM = 0.76

Study	Score	Screening Tool	Type of Study	Subjects	Raters	Video/Real Time	Analysis	Results
Minick et al (2010)	14/18 Good	FMS	Inter-rater	-40 healthy college students (23 female, 17 male) -Recruited by word of mouth -Mean age: 20.8	-4 raters, 2 expert 2 novice -Expert; developed the tool with 10+ years using the tool -Novice; taken standardised introductory course and less than one year's experience using the tool -Raters scored independently	**Video review** -No information on pausing or number of times video was reviewed given.	-Weighted kappa for each test between pairs of raters -Averaged scores of novice vs. expert raters compared with weighted kappa	-Novice raters: agreement; excellent for 6/17, substantial for 8/17 and moderate for 3/17 components. k = 0.53-1.0 -Expert raters: agreement; excellent 4/17, substantial for 9/17 and moderate for 4/17 components. k = 0.40-0.95 -Expert vs. novice: excellent for 14/17, substantial for 3/17 components. k = 0.74-1.0

Study	Score	Screening Tool	Type of Study	Subjects	Raters	Video/Real Time	Analysis	Results
Padua et al, 2009	15/19 Good	LESS	Inter-rater and Intra-rater	-50 subjects (25 males, 25 females), incoming freshman to military academies -Demographics of total population (n= 2691) subjects were recruited from - Height:172.9 ± 7.12 (males) 165.94 ± 6.63 (females) -Weight: 77.54 ± 12.34 (males) 63.12 ± 7.88 (females)	-Two raters -No other information given on raters -Raters scored independently	**Video** -Able to use; play, pause and rewind buttons	**Inter-rater** - ICC 2,k and SEM for overall score **Intra-rater** ICC 2,1 and SEM for overall score	**Inter-rater** -Good reliability; ICC = 0.84; SEM = 0.71 **Intra-rater** -Excellent reliability; ICC = 0.91; SEM = 0.42
Teyhen et al, 2012	15/19 Good	FMS	Inter-rater and intra-rater	-64 active-duty service members -Healthy adults -Gender not determined	-8 raters -All physical therapy students with same experience	**Real time** -Three trials of each test	-ICC (2,1), SEM and 95% CI for total score -Weighted kappa for individual	**Inter-rater** -Agreement; excellent for 1/7 ; k = 0.82, substantial for

Study	Score	Screening Tool	Type of Study	Subjects	Raters	Video/Real Time	Analysis	Results
Teyhen et al, 2012				-Age: 25.2 ± 3.8 years old -Height: 175.5 ± 9.6 cms -Weight: 77.5 ± 12.5kgs	-Same FMS training -Scored independently and simultaneously		tests	5/7; k = 0.67-0.77 and moderate for 1/7; k = 0.45 individual items -Substantial agreement for overall score; ICC = 0.76; SEM = 0.92 **Intra-rater** -Agreement; substantial for 5/7; k = 0.60-0.76, moderate for 1/7; k = 0.59 and poor for 1/7; k = 0.29 individual items. -Substantial agreement for overall score; ICC = 0.74; SEM = 0.98

Study	Score	Screening Tool	Type of Study	Subjects	Raters	Video/Real Time	Analysis	Results
Kritz et al (Forthcoming , 2014)	14/19 Fair	MCS	Inter-rater and Intra-rater	-3 elite athletes that demonstrated good (athlete 1), moderate (athlete 2) and poor (athlete 3) movement patterns. -One male BMX athlete and 2 female netball players -Athletes were randomly selected, no further information on recruitment given	-58 raters for inter-rater, 12 for intra-rater -41 strength and conditioning coaches and 17 physiotherapists -26 had 1-5 years and 32 had ≥ 6 years of professional experience -Rater recruitment not described -Previous experience with the MCS was not stated	**Video Review** -Raters were given 45 days to rate the MCS videos -No restrictions on how many times they were permitted to watch the videos	**Inter-rater** -Average percentage agreement for total MCS score and individual components **Intra-rater** -Kappa coefficient for overall MCS score	**Inter-rater** -Percentage of agreement for overall MCS score was substantial-almost perfect; k = 0.72-0.88 -Almost perfect agreement was seen for; 6/8 tests -Substantial agreement seen for 2/8 **Intra-rater** -Raters showed substantial to perfect agreement; k = 0.73-1.00
Plisky et al, 2009	14/19 Fair	SEBT: Y Balance	Inter-rater and Intra-rater	-15 male collegiate soccer players -Age 19.7 ± 0.81 years	-2 raters - Rater 1 physical therapy assistant and certified athletic trainer with 10	**Real time** -Three trials in each of the three reach directions	**Inter-rater** -ICC (2,1), SEM and 95% CI used to evaluate individual and composite reach	**Inter-rater** -Excellent agreement for all three directions and total score ICC = 0.97-1.0

Study	Score	Screening Tool	Type of Study	Subjects	Raters	Video/Real Time	Analysis	Results
Plisky et al, 2009					years of experience -Rater 2 physical therapist with 7 years of experience -No information given on training and experience with Y-Balance -Scored independently and simultaneously		distance scores for both left and right **Intra-rater** -ICC (3,1) SEM and 95% CI used to evaluate individual and composite reach distance scores	**Intra-rater** -Excellent agreement for all three directions and for total score ICC = 0.85-0.89.
Hertel et al, 2000	13/19 Fair	SEBT	Inter-rater and Intra-rater	-8 female and 8 males recreationally active, healthy young adult volunteers -Age: 21.3 ± 1.3 years old -Height: 171.2 ± 6.7cms -Weight: 70.3 ± 10kg	-2 examiners -No further information given on raters - Raters scored independently	**Real-time** -Subjects performed 2 bouts of the eight directions on each day	**Intra-rater** - ICC and SEM for each rater for the 3 trials in the 2 bouts on both days **Inter-rater** -ICC and SEM for the 6 trials of each test on both days for all 8 directions	**Inter-rater** -ICC = 0.35-0.84 on day one -ICC = 0.81-0.93 on day two **Intra-rater** -ICC = 0.78 – 0.96 on day one –ICC =0.85 and 0.96 on day two

Study	Score	Screening Tool	Type of Study	Subjects	Raters	Video/Real Time	Analysis	Results
Demura and Yamada, 2010	12/19 Fair	SEBT	Inter-rater and Intra-rater	-30 healthy young males -No information on recruitment - Age: 17.1 ± 2.5 years old - Height: 169.7 ± 3.2 cms -Weight: 68 ± 3.1 kg	-No information given on raters -Not stated if raters scored side by side or independently	**Real time** -Subjects performed each of the 8 reach directions 10 times with a minute rest between trials	ICC	**Inter-rater** -High reliability ICC = 0.99 **Intra-rater** -High reliability from the second to ninth trial in all directions ICC = 0.92-0.99

Studies are listed from highest to lowest overall methodological score and in alphabetical order for studies with the same methodological score. Abbreviations: CI = confidence interval, cm = centimetres, FMS = Function Movement Screen, ICC = intraclass coefficient, k = kappa coefficient, kg = kilograms, LESS = Landing Error Scoring System, MCS = Movement Competency Screen, SEBT = Star Excursion Balance Test, SEM = standard error of measurement.

2.3 Results

2.3.1 Screening tools

Seven different screening tools were utilised over the thirteen studies. The Functional Movement Screen (FMS) was investigated in four of the studies (Minick et al., 2010; Onate et al., 2012; Schneiders et al., 2011; Teyhen et al., 2012). The Landing Error Scoring System (LESS) was utilised in three studies (Onate et al., 2010; Padua et al., 2011; Padua et al., 2009). Two studies (Demura & Yamada, 2010; Hertel et al., 2000) investigated the rater reliability of the Star Excursion Balance Test (SEBT). Plisky et al (2009) utilised a modified SEBT; the Y-Balance. One study, Ekegren et al. (2009) looked at dynamic knee valgus. Dynamic posture during three specific dance movements; namely pliés in parallel, pliés turned out in first position and pliés into relevé from a parallel position, were assessed by Karim et al. (2011) based on the pilot 2006 Dance USA Annual Post-Hire Health Screen for Professional dancers. Kritz et al. (Forthcoming: 2014) utilised the Movement Competency Screen (MCS).

2.3.2 Subjects

There was variation between the studies in; subject selection, subject demographics and sample size. Sample size differed between three and 64, although only two studies (Ekegren et al., 2009; Onate et al., 2012) gave justification for sample size.

Three studies (Kritz et al., Forthcoming: 2014; Padua et al., 2011; Padua et al., 2009) did not include the age of their subjects in their demographical data. Average age ranged from 15 ± 1 year (Ekegren et al., 2009) to 25.2 ± 3.8 years (Teyhen et al., 2012). Average weight for subjects ranged from 60.51kg (Karim et al., 2011) to 85.13kg (Onate et al., 2012). Average height ranged between 164.08 (Karim et al., 2011) and 181.41cm (Onate et al., 2012). Weight and height was not reported in two studies (Minick et al., 2010; Plisky et al., 2009). Both genders were included in seven of the studies. Three studies recruited only females (Ekegren et al., 2009; Karim et al., 2011; Onate et al., 2010), two studies recruited only males (Demura & Yamada, 2010; Plisky et al., 2009) and one study did not report gender of subjects (Teyhen et al., 2012).

Five studies (Ekegren et al., 2009; Karim et al., 2011; Onate et al., 2010; Onate et al., 2012; Schneiders et al., 2011) employed a sample of convenience, five studies (Hertel et al., 2000; Minick et al., 2010; Padua et al., 2011; Padua et al., 2009; Teyhen et al., 2012) used volunteers as subjects and three studies (Demura & Yamada, 2010; Kritz et al., Forthcoming: 2014; Plisky et al., 2009) did not give any information on how recruitment was carried out.

Four studies recruited subjects from a specific sport. Ekegren et . (2009) and Onate et al. (2010) recruited female soccer players, whilst Plisky et al. (2009) recruited male soccer players and Karim et al. (2011) recruited female professional dancers. Three studies (Padua et al., 2011; Padua et al., 2009; Teyhen et al., 2012) recruited subjects from military academies.

2.3.3 Raters

The number, experience and training of raters varied between studies. Kritz et al. (Forthcoming: 2014) had fifty eight raters, all of whom participated in the inter-rater portion of the study and twelve of whom participated in the intra-rater portion of the study. Eight raters were used in one study (Teyhen et al., 2012), four raters were used in two studies (Karim et al., 2011; Minick et al., 2010), two studies (Ekegren et al., 2009; Padua et al., 2011) used three raters, six of the studies (Hertel et al., 2000; Onate et al., 2010; Onate et al., 2012; Padua et al., 2009; Plisky et al., 2006; Schneiders et al., 2011) used two raters and one study (Demura & Yamada, 2010) did not state the number of raters.

The experience and training of screening tool implementation was the same for all raters in four of the studies (Ekegren et al., 2009; Padua et al., 2011; Schneiders et al., 2011; Teyhen et al., 2012). Four of the studies (Karim et al., 2011; Minick et al., 2010; Onate et al., 2010; Onate et al., 2012) compared raters with varying amounts of experience with and training in the implementation of the screening tool of interest. The remaining five studies did not give any information on rater experience or training with the screening tool of interest.

To ensure that raters were blinded to the score of the other rater/s or to their previous score, raters in seven of the studies (Hertel et al., 2000; Karim et al., 2011; Onate et al., 2012; Padua et al., 2011; Plisky et al., 2009; Schneiders et al., 2011; Teyhen et al., 2012) scored simultaneously and independently. In contrast raters in five of the studies (Ekegren et al., 2009; Minick et al., 2010; Onate et al., 2010; Padua et al., 2009) watched and scored the movement screens independently. Although the raters in the Kritz et al. (Forthcoming: 2014) study scored independently, it was not made clear whether the raters involved in the intra-rater portion of the study were blinded to their previous score. No information was given on whether raters were blinded to the score of the other rater or their previous score in one study (Demura & Yamada, 2010).

2.3.4 Video/Real Time

In eight of the studies (Demura & Yamada, 2010; Hertel et al., 2000; Karim et al., 2011; Onate et al., 2012; Padua et al., 2011; Plisky et al., 2009; Schneiders et al., 2011; Teyhen et al., 2012) the screening tool of interest was scored in real-time. In the remaining five studies raters utilised video review to score the screening tool of interest.

2.3.5 Statistical Analysis

Five studies (Demura & Yamada, 2010; Hertel et al., 2000; Padua et al., 2011; Padua et al., 2009; Plisky et al., 2009) reported intraclass-coefficient (ICC) values with or without standard error of measurement (SEM). Only one study (Teyhen et al., 2012) calculated minimal detectable change score. Kappa coefficients alone were reported in Minick et al (2010). Four studies (Onate et al., 2010; Onate et al., 2012; Schneiders et al., 2011; Teyhen et al., 2012) reported both kappa and ICC values. Kritz et al. (Forthcoming: 2014) reported kappa coefficient for intra-rater agreement but used percentage of agreement for inter-rater agreement. Percentage of agreement alone was reported for Karim et al. (2011), whereas Ekegren et al. (2009) reported kappa coefficients along with percentage of agreement.

2.3.6 Study Findings

Inter-rater

Reliability for overall scores was generally higher than for individual items or tests. Of the thirteen studies that looked at inter-rater reliability, eleven studies (Demura & Yamada, 2010; Ekegren et al., 2009; Hertel et al., 2000; Kritz et al., Forthcoming: 2014; Onate et al., 2010; Onate et al., 2012; Padua et al., 2011; Padua et al., 2009; Plisky et al., 2009; Schneiders et al., 2011; Teyhen et al., 2012) reported reliability for the overall score and found reliability to be substantial to excellent. Scores for individual items or test for inter-rater reliability were on the whole higher than the intra-rater reliability scores. Of the eight studies that reported reliability statistics for individual items, three studies (Kritz et al., Forthcoming: 2014; Plisky et al., 2009; Schneiders et al., 2011) showed substantial to excellent reliability for all individual items. Moderate/good to excellent reliability was seen in three of the studies (Minick et al., 2010; Onate et al., 2010; Teyhen et al., 2012). The remaining two studies showed a much wider range of reliability. Karim et al. (2011) reported percentage of agreement as high (70-100%) for seven of their ten measurements. However the other three measurements resulted in percentage of agreement between 40-70%. Onate et al. (2012) reported kappa scores between 1.00 and 0.33 for the individual items of the FMS.

Intra-rater

Similar to the inter-rater results, intra-rater reliability was much higher for overall score of the screening tool of interest than for the individual items/tests of the screening tool. Of the eight studies (Demura & Yamada, 2010; Ekegren et al., 2009; Hertel et al., 2000; Kritz et al., Forthcoming: 2014; Onate et al., 2012; Padua et al., 2009; Plisky et al., 2009; Teyhen et al., 2012) that looked at intra-rater reliability, all reported moderate to excellent agreement, once learning was accounted for. Three of the studies reported reliability statistics for individual items or tests as well as for overall score. Onate et al. (2012) and Teyhen et al. (2012) showed excellent to poor and substantial to poor reliability respectively for the individual items of the FMS. Plisky et al. (2009) showed excellent reliability for the anterior, posteromedial and posterolateral reach distances.

2.4 Discussion

For a screening tool to be accepted as a method of accurately predicting injury risk, it must be shown to be a valid and reliable test. Although a measure can be shown to be reliable without being valid the reverse is not true (Hayen, Dennis, & Finch, 2007). It is therefore important to ascertain the reliability of screening tools prior to establishing their validity. For a screening tool to predict injury risk the accuracy of its measurements; overtime, by different raters and within the given subjects, is paramount (Hayen et al., 2007).

2.4.1 Screening tool

The premise behind pre-season screening tools is to be able to screen a large group of athletes to identify those that may be at an increased risk of injury, so that resources can be directed to those who may benefit more from them. To screen large groups of people, the screening tool will ideally; not utilise expensive and/or non-portable equipment, be able to be implemented and scored reliably by a variety of people, be time efficient and be cost effective.

All of the seven screening tools in this review were purported as screening tools that may be able to identify movement patterns that may increase injury risk. Although all these screening tools utilise a form of movement based testing and look to assess multiple domains of function, they are all distinctly different.

Star Excursion Balance Test

The SEBT measures reach distance across a series of directions whilst maintaining dynamic balance on the stance foot (Gribble, Hertel, & Plisky, 2012). Plisky et al. (2006) describe the SEBT as an inexpensive quick method of measuring functional balance via lower limb reach distance. The testing of dynamic balance rather than static balance is advantageous in the sporting population as, along with challenging the ability to remain upright and steady, it also encompasses the additional demands of proprioception, range of motion and strength (Gribble, 2003). The SEBT consists of a grid marked out on the floor with eight lines extending out at forty five degrees to each other. The lines are labelled for the direction of reach: anterior, anteromedial, anterolateral, medial, lateral, posterior, posteromedial and posterolateral. The subject places the foot to be tested on the centre of the grid and is then instructed to reach the opposite foot as far along the corresponding reach direction as they are able and to lightly touch the foot down. The subject must maintain their stance foot position and balance throughout the manoeuvre.

Y-Balance

The Y-Balance Test is a modified version of the SEBT. It was developed by Plisky (2009) to; 1) capture the greatest amount of information about dynamic balance in the shortest period of time, 2) improve repeatability of the SEBT and 3) to standardise test performance. Instructions for the SEBT mandate that each reach direction needs to be tested. Performing the eight reach directions bilaterally with sufficient practice trials to combat the learning effect is time consuming (Gribble, 2003; Hertel, Braham, Hale, & Olmsted-Kramer, 2006; Robinson & Gribble, 2008a). Recent studies have shown that there is redundancy in performing all eight reach directions (Hertel et al., 2006; Robinson & Gribble, 2008b). Due to the redundancy in performing all eight reach directions of the SEBT, the Y-Balance reduces the number of reach distances to only three.

The Y-Balance differs from the SEBT as it requires an instrumented device rather than laying tape on the ground. The device consists of a central foot plate and three pipes extending out with the posterior pipes positioned 135 degrees from the anterior pipes and 45 degrees from each other. The subject is required to push a plastic target along the required reach direction's pipe, which is marked in 5mm increments. Unlike the SEBT protocol, hands on hips are not mandated, stance foot is placed with the most distal aspect of the shoe at the start line and testing order is standardised. Trials are discarded or repeated if; 1) unilateral stance was not maintained, 2) foot contact with reach indicator was lost whilst the reach indictor was in motion, 3) the reach indicator was used for stance support and 4) the reach foot was not returned to the start position under control.

Dynamic Knee Valgus

Dynamic knee valgus was assessed by Ekegren et al. (2009). Subjects performed a drop-vertical jump (DVJ) task onto a force-plate embedded into the ground. Raters were to rater the individual as ‘high risk’ for ACL rupture if the patella moved inwards and ended up medial to the first toe or ‘low risk’ if the patella landed in line with the first toe upon landing of the DVJ.

Dynamic posture during three specific dance moves

The musculoskeletal screen that was used by Karim et al. (2011) was based upon the pilot 2006 Dance USA Annual Post Hire Health Screen for Professional Dancers. The screen was developed by the Dance/USA Taskforce on Dancer Health with the intention of developing a screen specific to dancers. It consists of one hundred and one items in the categories; static posture, Beighton 9-point hypermobility test, flexibility, strength and dynamic posture. The current review only assessed the dynamic posture part of the screen as it fitted the criteria of a movement based screen. The dynamic posture assessed three dance specific movements; plié in parallel, plié in turn out (hips externally rotated) and relevé to demi-pointe parallel balance. The rater assessed; anterior, posterior and lateral tilts of the pelvis, pelvic rotation, knee varus/valgus, foot pronation or supination and ankle inversion/eversion during the performance of each of the three dance specific movements.

Movement Competency Screen

The MCS was designed as a tool to determine safe training loads for athletes following an assessment of their movement competency (Kritz et al., Forthcoming: 2014). It was developed to help practitioners better understand an athlete’s movement competencies in fundamental movement patterns that are commonly used in activities of daily living, sport and sport specific training (Kritz et al., Forthcoming: 2014). The MCS consists of five tests (bodyweight squat, lunge and twist, bend and pull, push up and single leg squat) that assess eight fundamental movement patterns. Two of the tests, namely the lunge-and-twist and the bend-and-pull, assess two fundamental movement patterns within the same individual test. Raters assess movement quality for each body segment (head, shoulders, lumbar, hips, knees, ankles and feet) by check-marking the body segments that do not achieve an acceptable position, according to MCS screening criteria, whilst the athlete performs the movement test. The movement pattern is given a rating between one and three depending on the number of and which body segments are check-marked. Three of the eight fundamental movement patterns are assessed bilaterally and the lower of the two scores is recorded to account for asymmetry.

Functional Movement Screen

The FMS was designed to evaluate the body`s kinetic chain system, where the body is evaluated as a linked system of interdependent systems, which often work in a proximal to distal direction to initiate movement (Cook et al., 2006; Schneiders et al., 2011). The FMS looks at the concept of dynamic stability versus mobility (Kiesel & Honarbakhsh, 2009). It was developed as a screening tool to be performed on individuals without any recognised pathology (Kiesel & Honarbakhsh, 2009) and was designed to; 1) grade movement quality, 2) be body relative, 3) be time and space efficient and 4) to utilise minimal equipment and expense. The FMS comprises of seven tests (deep squat, hurdle step, in line lunge, shoulder mobility, active straight leg raise, trunk stability push up and rotary stability) that are graded from one to three. A clearing test is done for two of the tests, which are scored as zero if test positive. Five of the seven tests are assessed bilaterally and the lower of the two scores is recorded to account for asymmetry (Kiesel & Honarbakhsh, 2009).

Landing Error Scoring System

The LESS is an assessment tool that evaluates a dynamic sport-specific task (Padua et al., 2011). The LESS consists of seventeen items that are scored whilst the subject performs the DVJ. The subject jumps from a 30cm box to a distance fifty percent of their height and then perform a maximal vertical jump. The original LESS is scored by video review of three trials of the DVJ. Individual joint motions are scored at various specified moments in the landing sequence. Each component of the LESS is based on previous research that has been identified to increase risk of anterior cruciate ligament rupture (Onate et al, 2010). Of the three studies in the current review that looked at the LESS, only one followed the protocol of the original LESS (Padua et al., 2009). Onate et al. (2010) also utilised videotapes and described the original LESS, however their results would suggest that they only looked at 15 items rather than 17. No explanation is given of how many items they used and why, nor how each of the individual items are scored. Padua et al. (2011) developed a modified real time LESS RT-LESS) that assessed ten items rather than 17 and is performed and scored in real time.

Comparing screens

Of the seven screens included in the review, only the FMS and MCS look at both upper and lower body movement patterns. The FMS, MCS and Dynamic Posture during Dance Specific Movements are the only screens that target multiple movement patterns. The remaining screens focus instead on a specific skill, movement competency or target prevention of a specific injury.

Both the FMS and MCS are concerned with how the body moves as a whole, rather than focussing on isolated joints or skills. In the realm of injury prevention, this full body approach may well hold a key in identifying athletes that are at an increased injury risk. Not screening the body as a whole may fail to encompass how the kinetic chain, as a whole, moves to compensate for a certain deficiency. For example if the shoulder mobility test is limited this will also affect the subject's deep squat test in the FMS. Although the FMS encompasses all the fundamental movement patterns, aside from the pull pattern, it's focus is on mobility versus stability rather than pure movement patterning. The FMS encompasses fundamental movement patterns into tests that also assess; postural stability, core strength, flexibility and neuromuscular co-ordination, it also includes tests to capture elite performance by complex movement patterns requiring proper neuromuscular co-ordination and energy transfer from one segment of the body to another (Kiesel & Honarbakhsh, 2009). This contrasts with the MCS that is focused on how the kinetic chain performs pure fundamental movement patterns. It may be that these two screening tests can be used together to assess fundamental movement patterns and mobility versus stability. The utilisation of a screen such as the FMS or MCS is recommended if the movement and functioning of the kinetic chain as a whole are of interest. Although both these screens assess fundamental movement patterns, neither puts these patterns into a sport specific context. The addition of sport specific testing to these screens may increase their usefulness and relevance in the sport science realm.

The SEBT, for most subjects, represents a new complex movement task that they will not have been previously exposed to. Therefore a large learning effect can be expected and needs to be controlled for. When first performing a complex tasks such as the SEBT, variability in performance is to be expected which will only decrease with practice (Kinzey & Armstrong, 1998). Robinson and Gribble (2008a) and Munro and Herrington (2010) showed that four practice trials were required to attain stability for excursion distance of the reach leg angular displacement in all three planes of movement of the hip and knee of the stance leg.

The SEBT has been shown to be a valid tool in lower extremity injury, chronic ankle instability and ACL deficient knees (Herrington, Hatcher, Hatcher, & McNicholas, 2009; Olmsted, Carcia, Hertel, & Shultz, 2002; Plisky et al., 2006). Like the FMS the SEBT challenges a variety of skills namely; proprioception, range of motion and strength. This makes it a great screening tool for sports that require dynamic balance and that are known to leave athletes susceptible to chronic ankle instability and anterior cruciate ligament (ACL) deficiency. The short falls of the SEBT, when screening for global injury risk, are; 1) it does not assess fundamental movement patterns, 2) it does not assess upper body function and 3) it measures reach distance rather than movement quality. To truly understand how an athlete moves or functions as a whole, upper body patterns must also be taken into consideration.

Instead of screening for injury risk in general, like the FMS and MCS, some screening tools aim to screen for increased risk of sustaining a particular injury. Due to the four to six times greater risk of females sustaining an ACL rupture than their male counter-parts (Hewett et al., 2005), screens such as the LESS and dynamic knee valgus aim to identify factors that increase the risk of sustaining this injury. Being able to identify and screen for risk factors that may be associated with ACL injury is paramount in women's sports that require jumping, landing pivoting and deceleration skills, such as netball, to reduce the health care spending and long term health issues related to this injury.

The drop vertical jump (DVJ) has typically been used in studies as the jump-land test of choice to determine knee valgus and risk of ACL injury. Ford et al. (2003) reported a significant difference of 11 degrees between males and females for greater maximum valgus during the DVJ. The authors stated that this represents a key neuromuscular gender difference in what is a sports specific movement. Hewett et al. (2005) showed that female athletes who demonstrated increased dynamic valgus with the DVJ are at an increased risk of ACL rupture. By contrast the LESS, which also utilises the DVJ, was unable to predict ACL injury in a cohort of 5,047 high school or college athletes in a study undertaken by Smith et al. (2012). These results support the inclusion of a measure of dynamic knee valgus in movement based screens attempting to predict injury risk in female sports, such as netball, requiring the sport specific tasks of; deceleration, landing and/or pivoting manoeuvres. More research needs to be directed at the validity of reliable tests that measure dynamic knee valgus to identify which test is best to determine increased risk of ACL rupture.

Sport specific

Having a dynamic component of a screen, which is specific for a particular sport or the movement patterns and competencies required for that sport, is useful. Karim et al. (2011), in addition to all the battery of static tests they undertook, included a sports specific measure of dynamic posture during three ballet moves. Several of the postural components that were assessed whilst performing the ballet moves, were also assessed with static posture. Assessing these items whilst in motion and within a sport specific movement, may increase the ability to identify increased injury risk. The addition of sports specific tests to movement based screens may help to develop an understanding of how the athlete moves whilst participating in the sport and the specific forces and movements that their body is forced to withstand. It may also help to identify key deficits in movement competencies or patterns that may pre-dispose the athlete to injury that are not seen in traditional static or generic non-sport specific screening tools.

Screening tools that require expensive non-portable equipment were not included in this review. The real-time LESS reduces the need for video cameras and means to play back the recordings, which will decrease time and cost for LESS screening. Being able to score the screening tool of interest in real time will also increase the timing efficiency of the screening tool and make it a more viable option in a real world or clinical setting.

2.4.2 Subjects

Sample size justification

Justification of sample size was only given for two of the reviewed studies (Ekegren et al., 2009; Onate et al., 2012). Sample size justification is important as it can affect whether or not statistically significant results can be drawn from the study (Urdan, 2010). Future studies looking at the reliability of Movement Based Screens should justify their sample size.

Target population

When ascertaining the reliability of a screening tool selecting subjects that closely resemble the intended population in which the screening tool will be utilised is paramount. Most of the reviewed studies recruited young, healthy, physically active subjects. This is a valid strategy if this is the population that the screening tool will be utilised in. Both Ekegren et al. (2009) and Onate et al. (2010) targeted and recruited only females in their studies that looked at tools which established increased risk of ACL injury, due to the gender disparity of this injury. For screening tools that are intended for use in a variety of sports across different levels, it would be expected that the demographics of the target population would also be varied, meaning that the subjects in the sample population should reflect this. Similarly, if the screening tool is intended for use with recreational rather than elite athletes, recreational athletes should be targeted and vice versa. If this is not done, then caution should be applied when implementing the screen to a population that differs from the study population, as reliability of the screen may be affected by this population change. Ensuring an age appropriate sample population is also important owing to the maturation of movement patterns, changes in tissue and differing injury profiles seen with in different age groups (Brukner & Khan, 2012).

2.4.3 Raters

A screening tool with high inter-rater reliability using raters with varying levels of experience and training is advantageous in both research and real world clinical settings. Screening tools that require intensive training and high levels of experience to achieve high levels of inter-rater reliability make meaningful clinical implementation of the screening tool difficult.

The overall FMS score was shown to have excellent inter-rater agreement between raters of differing experience and training (Onate et al., 2012) and excellent (Schneiders et al., 2011) to moderate (Teyhen et al., 2012) agreement amongst raters with the same amount of experience and training. The decreased clinical experience of the physiotherapy students compared to the raters in the other two studies, may account for the decrease in ICC seen for Teyhen et al. (2012). Scores for individual tests of the FMS scored greater variability in agreement across the four studies looking at the FMS. Studies using raters with similar levels of training and experience showed substantial to excellent (Schneiders et al., 2011) and moderate to excellent (Teyhen et al., 2012) agreement for all FMS individual test scores and studies with raters involving differing experience and training showed substantial to excellent agreement for all tests (Minick et al., 2010; Onate et al., 2012), except for the hurdle step where poor agreement was seen in the Onate et al. (2010) study. Interestingly in the Minick et al. (2010) study, agreement between novice raters was better that agreement between the expert raters. Although variability is seen between the studies for levels of agreement between raters for individual tests, this does not appear to be affected strongly by rater experience or training.

Likewise rater experience does not appear to have any effect on the intra-rater reliability of the FMS. Onate et al. (2012) used a certified FMS specialist in their study whereas physical therapy students were used in the Teyhen et al. (2012) study. The results of the studies showed similar results. Both studies had poor to excellent agreement for individual items as well as substantial and excellent agreement for overall score being seen respectively in the Teyhen et al. (2012) and Onate et al. (2012) studies. One would assume that a certified FMS specialist would have shown better agreement with the individual items than the physical therapy students, although this was not seen.

The Kritz et al. (Forthcoming: 2014) study differed from the other studies included in the review due to the large number of raters and small number of subjects performing the MCS. The larger number of raters may increase variability of scores; however the small number of subjects performing the MCS may very well negate this variability (Urdan, 2010). Having a large number of raters from different professions and with differing lengths of professional experience is useful, as this may be a real world representation of who will be scoring the screening tool. Kritz et al. (Forthcoming: 2014) did not report rater experience or training with the MCS in the methodology, however the reader can assume, from the discussion, that the raters received the same training and had no previous experience with rating the MCS. As this was not made clear, it was not possible to ascertain if rater training and experience affected inter-rater reliability. We can however say that a large group of raters, with differing professional backgrounds and experience, had substantial to almost perfect agreement in total MCS score for the three athletes. This is a very useful finding as it suggests that the MCS can clinically be used as a screening tool by professionals from different professions and with varying levels of experience.

With respect to the SEBT it was not possible to ascertain if rater experience and training affects inter-rater reliability for the SEBT as no information was given about the training and experience of the raters in any of the studies (Demura & Yamada, 2010; Hertel et al., 2000; Plisky et al., 2009).

Of the three studies looking at the LESS only two, Padua et al. (2011) and Onate et al. (2010), gave information about rater experience and training. Excellent inter-rater reliability was seen between the expert and novice raters (Onate et al., 2010) and moderate to excellent reliability was seen between the raters from the same background (Padua et al., 2011). The Padua et al. (2011) study rated subjects in real time whereas the Onate etal. (2010) study rated subjects from video review, which may have had more of an effect on inter-rater reliability than rater experience and training.

Blinding raters to a previous score or to the score of the other rater/s is crucial for removing any risk of bias (Rumsey, 2011). The reviewed studies employed different methods in which to do this. Raters scored independently and simultaneously in six of the studies (Hertel et al., 2000; Karim et al., 2011; Onate et al., 2012; Padua et al., 2011; Schneiders et al., 2011; Teyhen et al., 2012). This methodology ensures that, especially in real time, the raters are scoring the same demonstration of the test. This methodology however may be affected by the position the raters are standing in to view the subject. Attempts are made to have raters view the subject from the same place; however this may have an effect on blinding if raters score side by side and are then able to view the score of the other rater.

2.4.4 Video/Real Time

Screening that is done in real time is more cost and time efficient than video review making it preferable in the real world setting when screening large groups. Due to time restraints placed on clinicians it is important that reliable and valid tests are developed that can be scored in real time (Padua et al., 2011). Having the ability to rewind, pause or replay allows raters another chance to review the movement. It may then be assumed that studies that are rated off video may have higher reliability than those done in real time.

When comparing studies looking at the FMS, three studies (Onate et al., 2012; Schneiders et al., 2011; Teyhen et al., 2012) were rated in real time whilst one (Minick et al., 2010) rated from video review. It would appear, when comparing these studies, that rating by video review was not overtly more reliable than rating in real time. The kappa scores for the Minick et al. (2010) study for individual items varied between 0.40-1.0, this was similar to the range of 0.33-1.00 seen by Onate et al. (2012). In contrast the kappa values for the Schneiders et al. (2011) and Teyhen et al. (2012) studies which were done in real time were between 0.73-1.0 and 0.45- 0.82 respectively, demonstrating higher kappa values than the video review from Minick et al. (2010).

When comparing studies that looked at the LESS, two studies (Onate et al., 2010; Padua et al., 2009) rated via video review and one (Padua et al., 2011) rated in real time. It was harder to compare these studies and determine whether rating via video or real time made any difference to results, as the LESS that was used in the Padua et al. (2011) study was adapted for use in real time scoring. The RT-LESS scores ten items rather than the seventeen scored in the LESS. However the two studies that scored via video review showed excellent agreement for overall score (ICC = 0.84-0.84) whereas agreement for real-time scoring was substantial to excellent (ICC = 0.72-0.81). It may be that for tests involving quick movements, such as jump and land, reliability is increased with video review.

2.4.5 Statistical Analysis

Seven of the studies (Ekegren et al., 2009; Kritz et al., Forthcoming: 2014; Minick et al., 2010; Onate et al., 2010; Onate et al., 2012; Schneiders et al., 2011; Teyhen et al., 2012) utilised Cohen's kappa in their statistical analysis.

Two studies (Ekegren et al., 2009; Karim et al., 2011) calculated percentage of agreement to ascertain reliability and Kritz et al. (Forthcoming: 2014) calculated average percentage of agreement to ascertain inter-rater reliability of overall MCS score and individual MCS components . Percentage of agreement is not considered to be a robust measure for reliability as it does not exclude agreement that may have occurred by chance (Riffenburgh, 2012). Results from these studies can therefore be expected to report higher agreement than if they had employed a different statistical method that took into account agreement that occurred by chance.

Eight of the studies employed ICC in their statistical analysis. Although ICCs are commonly used in reliability studies the methods used to calculate them vary between studies (Hayen et al., 2007) and to allow better comparisons between studies and to ensure that ICC is not inflated, the method which was utilised to calculate the ICC should be stated. SEM was reported by six of the reviewed studies (Hertel et al., 2000; Onate et al., 2012; Padua et al., 2011; Padua et al., 2009; Plisky et al., 2009; Teyhen et al., 2012) that used ICC to measure agreement. As some level of error is always present with continuous measurements, the inclusion of the SEM helps to determine the influence of measurement error on statistical results (Atkinson & Nevill, 1998) and helps determine real agreement rather than agreement that may have occurred by chance. Hayen et al. (2007) recommended that, along with ICC, details like SEM and variability of data, such as 95% CI, should also be reported to make comparing studies easier. The MDC is smallest change that indicates a real improvement for a group of subjects and represents the magnitude of change necessary to exceed the measurement error of two repeated measures at a specified confidence interval (Schwenk, Gogulla, Englert, Czempik, & Hauer, 2012). Teyhen et al (2012) calculated MDC with a 95% CI values of 2.1 (inter-rater) and 2.5 (intra-rater) points on the 21-point FMS scale. These results are promising and more studies need to utilise MDC scores to ascertain true improvement that is not due to a measurement error.

2.4.6 Study Findings

With the exception of Karim et al. (2011) all studies showed substantial to excellent agreement for inter and intra-rater reliability for all screening tools. For studies that reported agreement for both overall score and individual tests a trend was seen with agreement for overall scores being higher than that of individual tests. The smaller potential variation in scores for individual tests when compared to an overall score may account for the decreased reliability seen for the individual tests. The scoring system of the FMS is such that each individual test is scored out of three (excluding the two individual tests with clearing tests, that if positive will give a score of zero, which are scored out of four), meaning the subject can score either a one, two or three. Teyhen et al. (2012) found that no subjects scored zero on any of the clearing tests and only 18 of the 446 movement patterns assessed scored a one, meaning that the remaining 428 scored either a two or three. Likewise Schneider's et al. (2011) only scored a three for the rotary stability test for two subjects in their overall cohort of 209 subjects. This decreased variability in scores may account for poorer kappa scores seen for some of the individual FMS tests. Likewise the individual items of the MCS are scored out of three, decreasing the potential variation of scoring. As the range of scores in the Kritz et al. (Forthcoming: 2014) study is not given, it is not possible to tell if reliability is decreased further, by a reduced spread across scores. Of the three studies looking at the LESS, only Onate et al (2012), looked at agreement of individual tests as well as overall score. Percentage of agreement was generally high for all individual items (excluding overall impression), however kappa values varied between 0.46-1.0. Of note, knee flexion greater than 30 degrees had an agreement percentage of 90% but a kappa of 0.459. The

LESS individual items were scored on a binary Y/N basis for this study. The smaller number of possible responses may have decreased the kappa statistic for the individual tests in this study.

The variation in reliability of FMS individual tests may be associated with the less clearly defined descriptors of mid-range performance (Minick et al., 2010). When comparing individual tests the lowest agreement was seen for the; in-line lunge test (Minick et al., 2010; Schneiders et al., 2011; Teyhen et al., 2012), hurdle step test (Onate et al., 2012; Schneiders et al., 2011) and rotary stability test (Minick et al., 2010; Teyhen et al., 2012). Interestingly these three tests were identified by Minick et al. (2010) as the three with less clearly defined mid-range performance. To increase the reliability of the FMS overall score and individual tests, the scoring criteria for these tests, in particular, need to be more clearly defined. The author’s in the Kritz et al. (Forthcoming: 2014) study state that rater agreement was greater for the athlete with good movement patterns than it was for the athlete with fair and poor movement patterns. This may suggest that raters had trouble scoring the mid-range performances. The combined results from these two screens may suggest that whilst raters are able to agree on very good or very bad movement quality, the ability to rate mid-range quality is harder. Whatman, Hing and Hume (2012) investigated rater agreement for physiotherapists rating movement quality of four lower extremity functional tests (small knee bend, single leg squat, lunge and hop lunge). They found that both intra and inter rater reliability were increased when physiotherapists used a dichotomous scale rather than an ordinal scale to rate movement quality. This would suggest that screens that utilise a dichotomous method of scoring, asking the rater to simply ascertain yes/no may be more reliable than those that use an ordinal scale. Ekegren et al. (2009) simplified the rating of dynamic knee valgus, which may have been a factor in the higher levels of agreement reported. Raters had to classify the subject as being high risk if the patella moved inwards and ended up medial to the first toe and low risk if the patella landed in line with the first toe. The authors chose this methodology based on recommendations from a previous study (Chmielewski et al., 2007) that had used ambiguous rating criteria of good, fair and poor. Having raters focus on a single area of the body to rate a single criteria would increase reliability.

In contrast to the other screening tools, all studies that investigated the SEBT/Y-Balance showed excellent inter and intra-rater agreement for overall score and for individual reach distances. The only exception being; rater one on day one of the Hertel et al. (2000) study who had substantial intra-rater agreement (ICC = 0.78) for left lateral reach distance. The scoring for the SEBT/Y-Balance test differed from that of the other movement based tests as instead of movement quality being rated, distance achieved was recorded. The rater, therefore, was not concerned with how the subject moves to attain maximal reach distance, but simply how far the subject is able to reach. The rating of an objective measure such as reach distance may be a more reliable measure than rating movement quality.

Traditional scoring of the SEBT requires the subject's foot to remain in contact with the ground throughout the whole testing manoeuvre. This traditional scoring means that the rater is required to simultaneously focus on stance foot movement whilst attempting to mark reach distance. Plisky et al. (2009) allowed the subject's heel to lift off the ground to ensure the rater was focused on marking reach distance only. This change allowing the heel to lift off the ground may have accounted for the higher agreement seen in this study. To standardise scoring of the SEBT, this methodology will need to be taken up by all studies. Having a scale placed on the lines of the SEBT, which was seen in the methodology of the Demura and Yamada (2010) and Plisky et al. (2009) studies may also increase reliability of the test.

Percentages of agreement for the individual tests used in the Karim et al. (2011) study were varied. The fact that ten criteria needed to be marked for each movement may have decreased the agreement between the raters. The methodology also allowed raters to walk around the subjects whilst they were performing the movement tasks, meaning that each rater likely viewed each criterion from a different angle. Standardisation of the Dynamic Posture portion of the Dance specific screen is needed.

2.4.7 Summary

In general substantial to excellent reliability was seen for a variety of movement based screening tools. Reliability was better for overall scores than it was for individual tests in most of the screening tools reviewed. Rater training and experience did not appear to have a bearing on reliability of the screening tools. Although reliability was slightly better for some screening tools when they were scored off video review rather than in real-time, real-time analysis decreases monetary and time expenditure which is sought in a clinical setting. Standardised testing is important to ensure that meaningful comparisons can be made between studies and in the real world. Mid-range performance needs to be better quantified for individual tests in some screening tools.

2.5 Recommendations

- Screening tools should evaluate how the whole body moves and encompass tests that involve the upper and lower body

- Sport specific movement should be added to the screening tool

- If a certain sport is strongly associated with a particular injury, then tests that screen specifically for this injury should be included

- Screening in real time is more clinically applicable

- Subjects that closely resemble the target population should be recruited and described in methodology

- Raters need to be well described in methodology; it would appear that rater training and experience is not associated with better agreement

2.6 Conclusion

The literature review did not identify any screening tool, specific to netball that aimed to identify increased injury risk. Therefore a netball specific screen needs to be developed and the reliability of it established. From the review of the literature there are several key factors that need to be taken into account when developing and/or implementing a screening tool specific to netball. A movement based screening tool that assesses the entire kinetic chain including both upper and lower body movement patterns should be utilised. Sport specific tests that capture the sport specific skills such as; dynamic balance, jump and land and flexibility should be added to the screening tool. A measure of dynamic knee valgus whilst performing a jump and land manoeuvre should be added to help identify increased risk of ACL rupture, in what is a female dominated sport requiring, pivoting, jump and land and deceleration.

Chapter 3

Methodology

In this chapter the methodology of the study undertaken is explained. The Netball Movement Screening Tool (NMST) is introduced and the reasons why this screening tool was used are stated. Subject recruitment and ethical approval will then be explained. This is followed by the components of the NMST and how they are performed. The raters and how the rating was undertaken are explained and finally the statistical analyses employed are presented.

As no screening tools pertaining specifically to netball were identified in the literature review, the NMST was utilised in this study. The NMST was developed as a consequence of the literature review and in consultation with Netball New Zealand. The NMST includes; 1) MCS, to ascertain fundamental movement patterns, 2) Jump components to capture jump and land characteristics, 3) SEBT to capture the sport specific requirement of dynamic balance and 4) ASLR a component of the FMS to capture flexibility. The MCS was included to ascertain fundamental movement patterns. Whilst the MCS captures five basic movement patterns it does not have the sport specific jump or dynamic balance elements that are relevant to netball. Valgus loading of the knee during a jump-land manoeuvre has been shown to predict ACL risk (Hewett et al., 2005). As ACL injury has been shown as the most common netball injury (Flood & Harrison, 2009) the Jump components were added to the NMST. The SEBT has been shown to be a predictor of lower limb injury in high school basketball players (Plisky et al., 2006). As injuries sustained from netball occur predominantly in the lower limb (Flood & Harrison, 2009; Hopper & Elliot, 1993) and basketball and netball have a similar injury and skill profile (Flood & Harrison, 2009), the SEBT was added to the NMST. The ASLR was added to the NMST as flexibility has been shown to be one of the few variables that can be influenced in an injury prevention strategy (Meeuwisse, 1994). The purpose of this study was to ascertain the inter and intra-rater reliability of the NMST for screening in a cohort of adolescent female netball players.

3.1 Subjects

A sample of convenience of forty healthy 13-17 year old female high school secondary school netball players were recruited for the study. A convenience sample of secondary schools on Auckland's North Shore were contacted (via email and follow-up phone calls) and asked to participate in the study. Girls selected for the top four teams at each of the schools were invited to participate in this study. Subjects were excluded if they had sustained an injury in the past 6 weeks or experienced pain during any part of the movement screen. Twenty participants were assigned to the inter-rater group and twenty were assigned to the intra-rater group. Group assignment was determined by subject availability for follow-up testing.

Informed consent and assent were obtained from the participants and their guardians prior to data collection. Ethical approval was granted by Auckland University of Technology Ethics Committee (AUTEC 11/296). See Appendix A.

3.2 Data Collection

All participants were assigned a unique participant number. A pre-season questionnaire was completed by all participants to collect demographical data. Screening was undertaken by two physiotherapists who had undergone training in the delivery and scoring of the NMST patterns. This included consultation with the MCS creator, Matt Kritz, and in-house supervised practice in the delivery and scoring of the NMST. Rater one was a physiotherapist with 10 years of clinical experience and post-graduate Masters Degree and rater two was a final year physiotherapy student.

All participants were screened with the preseason Netball Movement Screening Tool (NMST) (Appendix D). The NMST consisted of four categories; 1) The Movement Competency Screen (MCS) category consisting of five individual tests; squat (Figure 2), lunge and twist (Figure 3), bend and pull (Figure 4), press up (Figure 5) and single leg squat (Figure 6); 2) Jump category; vertical jump A (Figure 7), vertical jump B (Figure 8) and broad jump (Figure 9); 3) Star Excursion Balance Test (SEBT) category (Figure 10) consisting of three reach directions; anterior, posterolateral and posteromedial and 4) Active straight leg raise category (ASLR) (Figure 11).

Participants were asked to wear shorts, tank tops and the shoes that they normally wore to play netball. An information sheet with instructions and photographs of the NMST was provided prior to testing and at the time of testing for review. Testing was undertaken as per the recommendations of the MCS developer (M.Kritz, personal communication, 17th June, 2011). Participants were given standardised verbal instructions prior to the performance of each test. Participants completed six repetitions of each of the MCS and Jump component individual tests, with three observed and scored from front on and three observed and scored from side on. The lunge, single leg squat, vertical jump B and broad jump were assessed bilaterally. For the bend and pull and lunge and twist patterns, where two patterns were scored within the same individual test, twelve repetitions of each were completed with six repetitions observed and scored from front on and six repetitions observed and scored from side on.

Figure 2: Squat. Showing start position (left) and lowering (right)

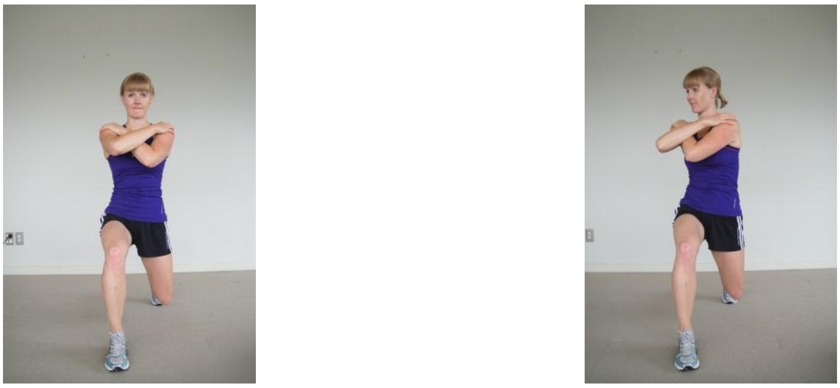

Figure 3: Showing lunge (left) and twist (right).

Figure 4: Bend and Pull. Showing start position (left), the bend (centre) and the pull (right)

Figure 5: Press up. Showing starting (left) and lowering (right) positions.

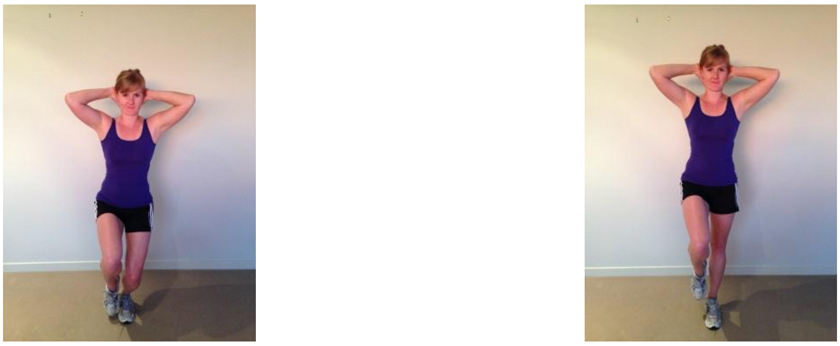

Figure 6: Single leg squat. Showing starting (left) and lowering (right) positions.

Figure 7: Vertical Jump A. Showing start (left), mid-flight (centre) and finish (right) positions.

Figure 8: Vertical Jump. Showing start (left), mid-flight (centre) and landing (right)

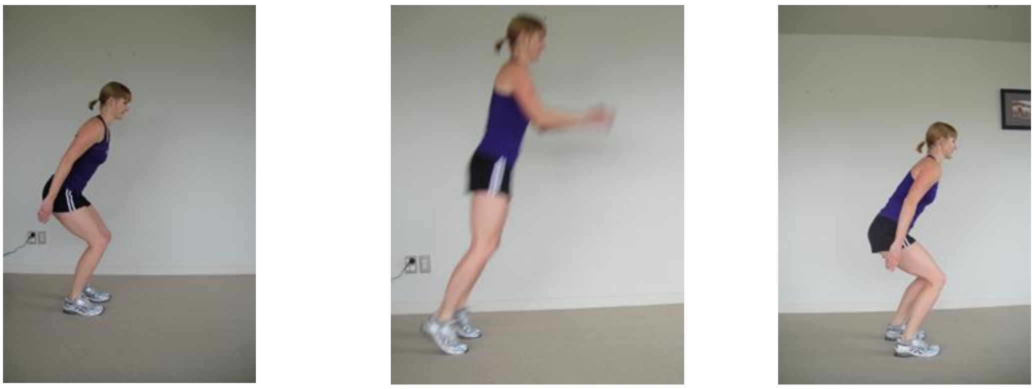

Figure 9: Broad Jump. Showing start (left), take-off (centre) and landing (right)

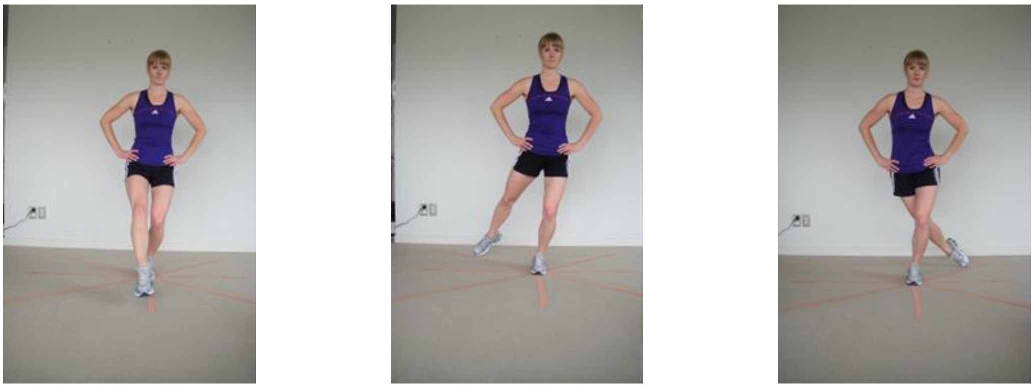

Figure 10: Star Excursion Balance Test. Showing the anterior (left), posterolateral (centre) and posteromedial (right) reach directions

Figure 11: Active Straight Leg Raise.

Figure 12: ASLR scoring. Showing score of one (left), two (centre) and three (right).

The MCS and jumping patterns were scored using a standardised form (See Appendix D). The scoring form lists primary and secondary areas of concern, namely: shoulders, lumbar spine, hips, ankles/feet, head, knees, depth and balance for each individual test. Optimal criteria for each test are seen in Appendix F. Any primary or secondary area of concern that was deemed to be sub-optimal at any time during any repetition of the individual test was given a check mark on the scoring form. For each individual test the number of sub-optimal check marks were added together to give a score out of three. A score of one was given when two or more primary regions and any secondary region were check marked. A score of two was given when one primary region and any secondary region were check marked. A score of three was given when no primary or secondary regions were check marked. For any patterns that were tested bilaterally the lower of the two scores was recorded. The overall maximum score for the MCS category is 21. The overall maximum score for Jump Category is nine.

The Star Excursion Balance Test (SEBT) was conducted in the same manner described by Plisky, Rauh, Kaminski and Underwood (2006). Participants were given the opportunity to review the instructional sheet (Appendix E) that was sent out with the participation sheet prior to beginning the SEBT trials. Participants were allowed six practice trials in each of the three directions to counter the significant learning effect seen in previous studies (Munro & Herrington, 2010; Robinson & Gribble, 2008a). Participants were instructed to assume the starting position by placing the arch of the foot being tested on the centre of the star and their hands on their hips. From the starting position participants were instructed to reach with their non-tested foot as far as they could along the reach direction line and to lightly touching down for three trials of each of the three reach directions. An erasable mark was drawn at the end of the big toe, by each rater, for each of the three trials for each reach direction and the greatest distance was measured with a measuring tape. Leg length (LL) was used to normalise reach distances for each subject as recommended by Gribble and Hertel (2003). This was measured with a measuring tape in supine from the anterior superior iliac spine (ASIS) to the lateral malleolus. Normalised SEBT reach distances were equated using the following formula as per Gribble and Hertel (2003):[(A + PL + PM)] / (LLx3) x 100

ASLR category was conducted in the same manner as Minick et al (2010). Participants were instructed to lay supine and to keep the non-test leg in contact with the ground throughout testing. Participants actively straightened their test leg and lifted it off the ground as far as they were able, without bending the test knee and ensuring that the non-test leg stayed in contact with the ground. Prior to testing the researcher marked two lines on the participant's leg with washable marker to divide the leg into three zones for scoring (Figure 12). One line was drawn at mid patella and the other line was drawn mid-thigh (half way between the ASIS and mid patella). Zone 1 was demarcated as distal to mid patella, Zone 2 between mid-patella and mid-thigh and Zone 3 superior to mid-thigh. The ASLR was scored by aligning a ski pole vertically with the medial malleolus of the test leg and recording which zone of the non-test leg the ski pole was situated in. A three was scored if the pole was situated in Zone 3, a two was scored if the pole was situated in Zone 2 and a one was scored if the pole was situated in Zone 1.

The scores of the MCS, Jump and ASLR categories of the NMST were calculated and combined to give an overall score. SEBT scores were not included in the overall NMST score, as the SEBT is scored on a continuous scale, whereas the other components of the NMST are scored on a dichotomous scale. The overall maximum score for the NMST is 33. Two raters simultaneously and independently (without consultation) scored the twenty subjects in the inter-rater group to examine inter-rater reliability. Instructions were given to all participants by the same rater for the inter-rater portion of the study. Rater one scored the twenty participants in the intra-rater group twice, with two weeks separating initial and follow up testing to examine intra-rater reliability. Raters received the same training and had the same experience in using the NMST.

3.3 Statistical Analysis

Data was analysed using SPSS statistical analysis software (version 19.0). Two-way mixed intra class correlation coefficients (ICC) for absolute agreement were calculated for; overall NMST, MCS, Jump, ASLR, SEBT (Left) and SEBT (Right) category scores. Standard error (SEM) of measurement was calculated using the equation: SE = Standard deviation (SD) / $\sqrt{n}$ as per Peate and Barton (2005). The minimal detectable change score (MDC) was calculated as SEM $\times 1.96 \times \sqrt{2}$ as per Schwenk et al (2012). Weighted Kappa values were calculated for individual NMST tests; squat, lunge, twist, bend, pull, push up, single leg squat, vertical jump A, vertical jump B and broad jump. ICC or kappa score over 0.80 was deemed to represent excellent agreement, 0.60-0.79 substantial agreement, 0.40-0.59 moderate agreement and below 0.40 poor agreement as per Minick et al. (2010), Schneiders et al. (2011) and Teyhen et al. (2012). Independent samples t-test ($p > 0.05$) was carried out to test for significant differences in demographics between the inter and intra-rater groups. Independent samples t-test ($p > 0.05$) was also carried out to test for differences between the mean scores of rater one and rater two (inter-rater group) and between initial and follow-up testing (intra-rater group) for MCS, Jump category and NMST.

Chapter 4

Results

Firstly in this chapter the demographics of the subjects are reported and compared between the inter and intra-rater groups. Then the inter and intra-rater agreement for the; NMST, MCS, Jump, ASLR and SEBT total scores and individual tests are reported.

4.1 Subjects

Forty female high school netball players, who met the inclusion criteria, participated in the study. Twenty of these participants underwent inter-rater testing and twenty participants underwent intra-rater testing. Mean and standard deviations of; age, height, weight, age started playing netball, number of years playing netball and number of teams currently playing for are given in Table 4. No significant difference was found between the two groups for any of the demographical or playing factors ($p > 0.05$).

Table 4: Mean and Standard Deviation of demographical data of all subjects in the inter rater and intra rater groups.

	All	Inter-rater	Intra-rater
Age	16.2 ± 1.1	16.1 ± 1.3	16.9 ± 0.9
Height	170.4 ± 7.7	168.7 ± 8.0	171.1 ± 7.4
Weight	63.3 ± 6.7	62.8 ± 8.0	63.8 ± 6.1
Age started Netball	7.8 ± 1.9	7.6 ±1.7	8.0 ± 2.1
Number of years playing	8.2 ± 1.8	8.3 ± 2.0	8.2 ± 1.6
Number of teams	1.6 ± 0.6	1.6 ± 0.6	1.5 ± 0.5

Independent samples t-test ($p < 0.05$).

4.2 Netball Movement Screening Tool

Excellent inter-rater agreement was found for total NMST score (ICC = 0.84; SEM = 0.25). ICC, SEM and MDC for overall category scores for inter-rater agreement are given in Table 5. Independent samples t-test ($p > 0.05$) found no significant difference between the mean scores of rater one and rater two for total NMST score in the inter-rater group. Mean NMST score for the inter-rater group was 21.05 ± 2.04 for rater one and 20.7 ± 2.11 for rater two. Mean and standard deviation (SD) for overall category scores for the inter-rater group are given in Table 6. Excellent intra-rater agreement was found for total NMST score (ICC = 0.96; SEM = 0.13). ICC, SEM and MDC for overall category scores for intra-rater agreement are given in Table 7. Independent samples t-test ($p > 0.05$) found no significant difference between the mean NMST scores at initial testing and follow up testing for the intra-rater group. Mean NMST score was 21.05 ± 2.16 at initial testing and 21.2 ± 2.09 at follow up testing. Mean and standard deviation for overall category scores for the intra-rater group are given in Table 8.

Table 5: Intraclass correlation for inter-rater agreement of overall scores for NMST categories.

	ICC	95% CI		SEM	95% CI		MDC
		Lower	Upper		Lower	Upper	
NMST	0.84	0.65	0.93	0.25	-0.18	0.88	13.8
MCS	0.77	0.49	0.91	0.18	0.16	0.78	7.5
Jump	0.65	0.29	0.84	0.21	-0.46	0.46	8.8
ASLR	1	-	-	1	-	-	1.7
SEBT (left)	0.99	0.98	1.0	0.26	-1.12	-0.07	28
SEBT (right)	0.99	0.99	1.0	0.20	-0.39	0.43	6.3

Two-way mixed effects model where people and effects are random and measures effects are fixed using absolute agreement coefficient.

Table 6: Mean and SD for overall category scores for the inter-rater group as scored by rater one and rater two.

	Rater One	Rater Two
NMST	21.05 ± 2.04	20.7 ± 2.11
MCS	13.75 ± 1.25	13.35 ± 1.35
SEBT (left)	87.62 ± 10.29	88.19 ± 10.29
SEBT (right)	88.48 ± 8.9	88.01 ± 8.91

Table 7: Intraclass correlation for intra-rater agreement of overall scores for NMST categories.

	ICC	95% CI Lower	95% CI Upper	SEM	95% CI Lower	95% CI Upper	MDC
NMST	0.96	0.91	0.98	0.13	-0.42	0.12	10.5
MCS	0.88	0.71	0.95	0.14	-0.23	0.33	3.2
Jump	0.88	0.71	0.95	0.16	-0.59	0.09	3.0
ASLR	0.90	0.77	0.96	0.05	-0.05	0.15	1.1
SEBT (left)	0.81	0.53	0.92	1.18	-5.35	-0.39	20.6
SEBT (right)	0.79	0.54	0.91	1.34	-4.81	0.78	23.6

Two-way mixed effects model where people and effects are random and measures effects are fixed using absolute agreement coefficient.

4.3 Movement Competency Screen

Substantial inter-rater agreement was seen for overall MCS score (ICC = 0.77; SEM = 0.18). ICC, SEM and MDC for overall category scores for the inter-rater group are given in Table 5. Independent samples t-test ($p > 0.05$) found no significant difference between the mean scores of rater one and rater two for MCS total score in the inter-rater group. Mean MCS score in the inter-rater group was 13.75 ± 1.25 for rater one and 13.35 ± 1.35 for rater two. Mean and SD for overall category scores for the inter-rater group are given in Table 6.

Excellent intra-rater agreement was seen for the overall MCS score (ICC = 0.88; SEM = 0.14). ICC, SEM and MDC for overall category scores for the intra-rater group are given in Table 7. Independent samples t-test ($p > 0.05$) found no significant difference between the mean MCS score between initial testing and follow up testing in the intra-rater group. Mean MCS score of 13.6 ± 1.31 was attained at initial testing and mean MCS score of 13.55 ± 1.05 was attained at follow up testing for the intra-rater group. Mean and SD for overall category scores for the intra-rater group are given in Table 8.

Table 8: Mean and SD for overall category scores for the intra-rater group for initial and follow up testing

	Initial testing	Follow up testing
NMST	21.05 ± 2.16	21.2 ± 2.09
MCS	13.6 ± 1.31	13.55 ± 1.05
SEBT (left)	89.80 ± 9.78	92.67 ± 8.82
SEBT (right)	91.17 ± 8.77	93.19 ± 9.98

Kappa values of individual MCS test scores ranged from 0.32-0.75 in the inter-rater group. Substantial inter-rater agreement was seen for the; squat, lunge and twist. Moderate inter-rater agreement was seen for the; bend, pull and single leg squat. Poor inter-rater agreement was seen for the push-up. Kappa data for individual MCS and Jump tests are given in Table 9. For the inter-rater group neither rater gave any subject a score of one for the lunge. Likewise neither rater gave a score of three to any subject for the push up and for the single leg squat. For the push up, rater one awarded a score of one to eighteen of the twenty subjects and rater two awarded a score of one to seventeen of the twenty subjects.

Table 9: Weighted kappa values for inter-rater agreement of individual test scores.

Individual Test	Kappa
Squat	0.61
Lunge	0.70
Twist	0.75
Bend	0.44
Pull	0.58
Push up	0.32
Single leg squat	0.57
Vertical jump A	0.66
Vertical jump B	0.62
Broad Jump	0.70

Kappa values of individual MCS tests in the intra-rater group ranged from 0.27-0.75. Substantial intra-rater agreement was seen for the; squat, lunge and push-up. Moderate intra-rater agreement was seen for the bend, pull and single leg squat. Poor agreement was seen for the twist. Kappa data for individual MCS and Jump tests are given in Table 10. No subjects received a score of one for the twist or the pull at either testing session. In the follow up testing session, no subjects received a score of one for the twist, whereas one subjects received a score of one in the initial testing session. At both initial testing and follow up, no subjects received a score of three for the push up or single leg squat.

Table 10: Weighted kappa values for intra-rater agreement of individual test scores.

Individual Test	Kappa
Squat	0.75
Lunge	0.71
Twist	0.27
Bend	0.66
Pull	0.57
Push up	0.62
Single leg squat	0.57
Vertical jump A	0.69
Vertical jump B	0.54
Broad Jump	0.77

4.4 Jump Category

Substantial inter-rater agreement was found for the overall jump category score (ICC = 0.65; SEM = 0.21). Excellent intra-rater agreement was seen for the overall jump category score (ICC = 0.88; SEM = 0.16). Kappa values of individual jump test scores in the inter-rater group ranged from 0.62-0.70. Substantial inter-rater agreement was seen for all three jump tests. Kappa data for individual MCS and Jump test inter-rater agreement are given in Table 9. Kappa values of individual jump test scores in the intra-rater group ranged from 0.54-0.77. Substantial intra-rater agreement was seen for the vertical jump A and the broad jump. Moderate intra-rater agreement was seen for the vertical jump B. Kappa data for individual MCS and Jump test intra-rater agreement are given in Table 10. At both initial testing and follow up no subjects received a score of three for the broad jump.

4.5 Star Excursion Balance Test

Excellent inter-rater agreement was seen for the SEBT left (ICC = 0.99; SEM = 0.26) and SEBT right (ICC = 0.99; SEM = 0.18) category scores. ICC and SEM for overall scores in the inter-rater group are given in Table 5. Independent samples t-test ($p > 0.05$) found no significant difference between the mean scores of rater one and rater two for the SEBT (right) and SEBT (left) in the inter-rater group. Mean SEBT (left) score for rater one was 87.62 ± 10.29 and mean score for rater two was 88.19 ± 10.29 in the inter-rater group. Mean SEBT (right) score for rater one in the inter-rater group was 88.48 ± 8.9 and mean score for rater two in the inter-rater group was 88.01 ± 8.91. Mean and SD for overall category scores for the inter-rater group can be given in Table 6.

Excellent intra-rater agreement was seen for the overall category scores of the SEBT (left) (ICC = 0.81; SEM = 1.18). Substantial intra-rater agreement was found for the SEBT (right) category scores (ICC = 0.79; SEM = 1.34). ICC and SEM for overall scores for the intra-rater group are given in Table 7. Independent samples t-test ($p > 0.05$) found no significant difference between the mean scores at initial testing and follow up testing for the SEBT (right) and SEBT (left) in the intra-rater group. Mean SEBT (left) score for the intra-rater group was 89.80 ± 9.78 at initial testing and 92.67 ± 8.82 at follow up testing. Mean SEBT (right) score for the intra-rater group was 91.17 ± 8.77 at initial testing and 93.19 ± 9.98 at follow up. Mean and SD for overall category scores for the intra-rater group are given in Table 8.

4.6 Active Straight Leg Raise

Perfect inter-rater agreement was seen for the ASLR category score (ICC = 1.0; SEM = 0). ICC and SEM for overall scores for the inter-rater group are given in Table 5. Excellent intra-rater agreement was seen for the ASLR category score (ICC = 0.90; SEM = 0.05). ICC and SEM for overall scores of the intra-rater group are given in Table 8.

Chapter 5

Discussion

Firstly in this chapter an overview of the purpose and results of the study is given. Next the inter and intra-rater agreement found in the NMST tool is discussed and compared to other studies looking at movement based screening tools that aim to predict injury risk. The inter and intra-rater agreement found in the current study for the MCS as a stand-alone screen is then discussed. The inclusion of the Jump tests in the NMST is then discussed and compared to the jump and land tests utilised in other movement based screens. How the SEBT and ASLR fit into the NMST is then discussed. Finally the limitations of the current study are stated.

5.1 Overview

The purpose of this study was to assess the reliability of a tool, designed to assess movement competency, in adolescent female netball players. The results of the study demonstrate excellent inter and intra-rater agreement for the NMST when screening adolescent female netball players. The results seen in the current study are similar to previous studies investigating the inter and intra-rater reliability of other movement screening tools. In the current study reliability for overall NMST and for the total category scores of the; MCS, Jump, ASLR and SEBT were shown to be substantial to perfect for inter-rater reliability and substantial to excellent for intra-rater reliability.

5.2 Netball Movement Screening Test

Agreement was substantial to excellent for overall and category scores for the NMST in the current study. However, individual tests showed more variable agreement and reliability. Individual tests of the MCS and Jump categories showed poor-substantial inter and intra-rater reliability. Onate et al. (2010), Onate et al. (2012), Schneiders et al. (2011) and Teyhen et al. (2012) also showed a similar pattern of higher reliability for overall score when compared to individual tests. Onate et al. (2010) showed excellent inter-rater reliability for overall LESS score and moderate to excellent inter-rater reliability for individual items. Similarly, Onate et al. (2012) showed excellent agreement for overall FMS score and poor to excellent agreement for individual items for both inter and intra-rater reliability. Schneiders et al. (2011) demonstrated excellent inter-rater reliability for overall FMS score and substantial to excellent inter-rater reliability for individual items of the FMS. Teyhen et al. (2012) showed substantial agreement for inter and intra-rater reliability for overall score of the FMS and poor to substantial intra-rater agreement and moderate to excellent inter-rater agreement was found for the individual items of the FMS. The current study showed the same pattern with agreement for individual tests being lower than agreement for overall category and NMST score.

In the current study, weighted kappa scores for inter-rater reliability showed moderate agreement for the bend, pull and single squat and poor agreement for the push up. For the intra-rater reliability, weighted kappa scores showed poor agreement for the twist and moderate agreement for the; bend, pull, single leg squat and vertical jump B. A high percentage of agreement can result in a low kappa value, known as the paradox of the kappa statistic, due to the sensitivity of kappa to prevalence (Feinstein & Cicchetti, 2000). The lack of spread across the three scores in some of the individual tests of the NMST may have under estimated rater agreement, as the kappa statistic is best utilised when prevalence is close to 50% (Gwet, 2008) . No subjects in the current study scored a three for the push up or the single leg squat, three of the forty subjects scored a three for the vertical jump B and only one subject scored a three for the broad jump. The reporting of agreement percentage as well as kappa may have been useful in the current study to counter the under estimation of rater agreement.

The minimal detectable change scores (MDC) for the NMST are not so promising. As stated earlier, the MDC is smallest change that indicates a real improvement for a group of subjects and represents the magnitude of change necessary to exceed the measurement error of two repeated measures at a specified confidence interval (Schwenk et al., 2012). For the NMST total score MDC score is 13.4/33 for the inter-rater group and 10.5/33 for in the intra-rater group. The reason for the high MDC scores seen in the current study may indicate higher levels of variability in the individual tests that make up the overall scores. Only one of the reviewed studies (Teyhen et al., 2012) calculated MDC scores and they found MDC95 values of 2.1 (inter-rater) and 2.5 (intra-rater) points on the 21-point FMS scale. These scores are substantially better than the current study and may indicate less variability in individual test scores than for the NMST. Further work in reducing the variability in the scoring will be required in future for the detectable change scores to be more meaningful in these tests.

Having a more structured approach to scoring each of the individual tests of the NMST, as is seen in the LESS, may increase inter and intra-rater reliability of this screening tool. This would ensure that all aspects of the movement pattern are being assessed every time the athlete is tested, rather than assuming that the rater assessed every part of the movement pattern. The LESS dictates which body parts are scored on each repetition and from which view these are to be viewed. In the current study it would appear that the single leg squat has the lowest reliability over both measures of reliability. Guidelines on which body parts are to be scored on each repetition and the view from which they are to be scored from, may increase agreement of the single leg squat in particular and of the individual tests of the NMST.

The current study found that the; NMST, MCS, Jump, ASLR and SEBT categories can be reliably administered to adolescent female netball players by raters with varying levels of clinical experience and the same training in delivery of the NMST. The ability for a screening tool to be administered by raters of varying levels of clinical experience is important and applicable to the real world setting. For the utilisation of a screen to become widespread, raters need to be trained to reliably deliver the screening tool without a great time and monetary expenditure. A screen that requires years of training and experience to be administered reliably is not as clinically applicable or useful when compared to one which can be applied reliably by raters with little experience or training. The NMST was shown to have excellent inter-rater agreement between a physiotherapy student and a physiotherapist with Post-Graduate qualifications and over 10 years of clinical experience. It is important to note however, that the raters in the current study underwent the same amount of training in the implementation of the NMST.

The addition of skill, sport, injury and gender specific tests to a movement based screening test is warranted. The ASLR, SEBT and Jump categories were added to the

NMST to capture the sport specific requirements for netball players. Although these were all added for differing reasons in an attempt to capture increased injury risk, it could be argued only the jump components score true movement quality. Although both the SEBT and ASLR are movement based, their scores are indicative of a measured value rather than movement quality. If movement quality is what we are trying to measure, the addition of the ASLR and SEBT is questionable. Validity studies need to be undertaken to ascertain which NMST tests, or combination thereof, are able to better capture increased injury risk.

The NMST, like many other screening tools, is aimed at screening only. It seeks to identify athletes with faulty movement patterns rather than understand why they demonstrate faulty movement patterns. The premise behind the utilisation of screening tests is to save time, money and resources, and to streamline these into athletes who may need them. If faulty movement patterns are demonstrated by the screening tool, a full musculoskeletal examination should be undertaken to further investigate what may be the cause of their dysfunction.

5.3 Movement Competency Screen

The MCS as a stand-alone screen was shown by the current study to have substantial inter-rater and excellent intra-rater agreement for overall MCS score. To date there has only been one other study that has investigated the reliability of the MCS. Kritz et al. (Forthcoming: 2014) demonstrated substantial to excellent inter rater reliability and substantial to perfect intra rater reliability for the MCS. Direct comparisons between the present study and the Kritz et al. (Forthcoming: 2014) study are difficult to make due to the differences in study design. Kritz et al. (Forthcoming: 2014) utilised 58 raters who scored three subjects deemed to demonstrate; good, moderate and poor movement patterns. Whereas the current study had two raters score 20 subjects with undetermined movement pattern ratings. Study designs aside, the combined results of these two studies indicate that the MCS can be considered to be a reliable screening tool.

The MCS was designed as a tool to determine safe training loads for athletes following an assessment of their movement competency. Athletes that demonstrated sub-optimal movement patterns were not loaded in the strength and conditioning environment until the causes of their faulty movement pattern were addressed and correct movement pattern could be demonstrated (M. Kritz, personal communication, 17 June, 2011). The MCS focuses on fundamental movement patterns, rather than focussing on mobility versus stability, which is the focus of the FMS. This focus on fundamental movement patterns caught the interest of other sports medicine practitioners and it was theorised that the MCS may be able to identify athletes with a higher risk of sustaining injury (M. Kritz, personal communication, 17 June, 2011). The validity of the MCS in identifying increased injury risk is yet to be ascertained. Future studies that compare the ability of the MCS as a stand-alone screen and the MCS as a part of a more sports specific screening tool, such as the NMST, need to be undertaken. The addition of sport, injury and gender specific tests to screening tools is thought to increase the ability of the screen to identify increased injury risk.

The ability for the adolescent female population of the current study to reach the benchmarks that are needed to achieve a score of three in the MCS for the push up and single leg squat is questioned in the current study. No subjects were given a score of three for either of these tests. This inability to achieve a score of three may indeed signal a predisposition for injury to this age group of female athletes, or calls for a change in the benchmark criteria for scoring a three for this test in this, and potentially other, subject populations.

Individual tests such as the bend and pull and lunge and twist that test two movement patterns simultaneously need good guidelines for testing to ensure that both movement patterns are properly assessed. When scoring the lunge and twist rater one check marked ‘knee’ for the lunge when the knee drifted medially as the subject twisted, whereas rater two check marked ‘knee’ for the twist for the same movement anomaly. The check marking for the same movement anomaly for different movement pattern may account for some of the variation seen between reliability for individual item scores compared with reliability of overall scores. More definitive guidelines may be needed for the scoring of these individual tests that simultaneously test two movement patterns.

5.4 Jump Category

As ACL injuries are common in female netball players (Flood & Harrison, 2009) the Jump Components were included in the NMST to capture the jump and land characteristics which may be demonstrative of increased ACL rupture risk. The current study found substantial inter-rater agreement and moderate to substantial intra-rater agreement for the Jump components. This result contrasts with the Onate et al. (2010), Padua et al. (2009) and Padua et al. (2011) studies that showed excellent, substantial and excellent to substantial inter-rater agreement respectively for the LESS. The LESS is more structured in the way that it is scored than the Jump components of the NMST. Padua et al. (2011) mandates where the rater should stand and from which number jump each item is scored. Stance width, maximum foot rotation position and initial foot-contact symmetry are all scored during the first repetition and are viewed from the front. Maximum knee valgus and amount of lateral trunk flexion are scored in the second repetition and are also viewed from the front. Initial landing of the feet and amount of knee flexion displacement are viewed from the side on the third repetition and amount of trunk flexion displacement is viewed from the side and scored on the fourth repetition. Total joint displacement in the sagittal plane and overall impressions are scored from all repetitions. This more structured approach may increase inter and intra-rater reliability and a similar approach should be undertaken for the Jump components of the NMST.

It is not known whether undertaking the three different jump manoeuvres in the NMST is a better representation of increased injury risk than the drop vertical jump that is used in the LESS. Ford, Myer & Hewett (2003) reported a significant difference of 11 degrees between males and females for greater maximum valgus during the drop vertical jump. The authors stated that this represented a key neuromuscular gender difference in what is a sports specific movement. Whether this same gender difference is observed in the three jump tests included in the NMST is unknown. Future studies comparing the validity of both the DVJ and Jump Tests included in the NMST are warranted. Due to the increased cost and disability that can come from ACL rupture and the frequency of ACL rupture within netball, it may also be suggested that the inclusion of the DVJ and/or LESS in the NMST, instead of the three Jump tests that are currently used, is warranted.

The biggest disparity between inter and intra-rater agreement was seen with the Jump overall score. Inter-rater agreement was substantial (ICC: 0.65) whereas intra-rater agreement was excellent (ICC 0.81). The ability to understand what movement patterns or anomalies the jump tests are screening for, at speed, may be increased with increased clinical experience (Whatman et al., 2012). As both raters viewed and scored the same performance of the three jump manoeuvres, rater positioning whilst scoring may also be a factor. The use of video for screening jump tests is relevant here as it would allow raters to get the same view of the subject and give them the ability to slow down or pause during the manoeuvre. The speed at which the jump tests occur, in real time, may be too fast for the naked eye to process, and this may affect rater agreement. Scoring in real time also does not allow for the option to pause or slow down movements, which may allow for more accurate scoring and may improve rater agreement.

5.5 Star Excursion Balance Test

In the current study the SEBT showed excellent and substantial to excellent reliability for inter and intra rater agreement respectively. These findings are similar to previous studies that have also shown excellent inter rater agreement and substantial to excellent intra rater agreement for the SEBT and the Y-Balance tests (Hertel et al., 2000; Plisky et al., 2009). The excellent inter and intra rater agreement of the SEBT make it a reliable screening tool. In the current study, over 50 percent of the screening time was taken up by the SEBT, due to the three practice trials in each reach direction followed by the three trials in each reach direction. Owing to the additional time component of the SEBT, studies looking at validity need to be undertaken to ascertain whether or not the SEBT is a valid tool to predict injury in the targeted population and therefore whether it is needed in the NMST.

As the SEBT is a complex movement task, variations in performance of subjects who are unfamiliar with the test can be expected (Kinzey & Armstrong, 1998). A learning effect that can only be abated by practice and repetition has been well documented for the SEBT (Kinzey & Armstrong, 1998; Munro & Herrington, 2010; Robinson & Gribble, 2008a, 2008b). Robinson and Gribble (2008a) and Munro and Herrington (2010) showed that, with the exception of the lateral reach direction, only three practice trials were needed to counter the learning effect of the SEBT. The results of the present study confirm the recommendation of three practice trials, as no significant difference in mean SEBT left or SEBT right scores were seen between initial testing and follow up testing. This finding adds weight to the body of research advocating the implementation of three practice trials for the SEBT.

Plisky et al. (2006) showed the anterior, posteromedial and posterolateral reach directions of the SEBT to be predictive of lower extremity injury in high school basketball players. Owing to the similarities between the sports of netball and basketball and their injury profiles found by Flood and Harrison (2009), it stands to reason that the SEBT may be a valid predictor of increased injury risk in netball players. Research to ascertain this is needed. The SEBT was also shown by Olmsted et al. (2002) to be efficacious in detecting reach deficits in subjects with chronic ankle instability when compared to matched controls without chronic ankle instability. However whether or not chronic ankle instability increases injury risk in netball players has not been validated. Similarly, Herrington et al. (2009) showed significant differences in the anterior, lateral, posteromedial, and medial reach directions between subjects with and without ACL deficiency. Significant differences were also seen for the medial and lateral reach directions between the controls and the un-injured limb of the ACL deficient subjects. Interestingly no significant differences in reach distance were seen between the injured and un-injured limb of the ACL deficient subjects. This may mean that subjects with decreased medial and lateral reach distances are predisposed to ACL injury, or that the ACL injury has caused a bilateral neuromuscular deficit in this patient population. Again studies need to be undertaken to ascertain the validity of the SEBT in detecting increased injury risk within the specific population of netball players.

The current study chose to only measure three of the eight SEBT reach directions as previous studies have shown redundancy in performing all eight reach directions (Hertel et al., 2006; Robinson & Gribble, 2008a, 2008b). The reach directions of; anterior, posteromedial and posterolateral, or essentially the Y Balance Test, were chosen for the current study as these three reach directions were shown to be predictive of lower limb injury in high school basketball players by Plisky et al. (2006). Whether or not these reach directions are more predictive for indicting increased injury risk in this population group is unknown.

The SEBT measures excursion distance and was designed as a tool for assessing dynamic balance. As a screening tool the SEBT does not quantify or score movement quality, rather it measures reach distance attained. Further validity studies should be undertaken to ascertain whether the full NMST, the NMST without the SEBT, or the SEBT alone are more predictive of injury risk in netball players

5.6 Active Straight Leg Raise

The ASLR assesses the active flexibility of the hamstrings and calf muscle complex, whilst maintaining a stable pelvis and active extension of the opposite leg (Cook et al., 2006). The ASLR is included in the NMST as decreased flexibility has been linked to lower limb injury (Davis Hammonds, Laudner, McCaw, & McLoda, 2012; Hartig & Henderson, 1999; O'Sullivan, McAuliffe, & DeBurca, 2012; Weldon & Hill, 2003) and flexibility has been shown to be one of the few variables that can be influenced in an injury prevention strategy (Meeuwisse, 1994). The ASLR was shown to have perfect inter-rater and excellent intra-rater agreement in the current study. The reliability of the ASLR has also been shown to be excellent in other studies. Minick et al. (2010) found excellent inter-rater agreement for the ASLR between; novice raters, expert raters and for the average scores of novice and expert raters. Likewise Schneiders et al. (2011) found excellent inter-rater agreement for the ASLR. Although Onate et al. (2012) demonstrated high kappa values for intra-rater agreement for the ASLR (kappa 0.88), the intra-rater agreement had a kappa value of 0.69, which was classified as poor agreement by the authors. When first administering the ASLR, the implementation and scoring of the test takes more than a cursory glance. However the rater involved in the intra-rater portion of the Onate et al. (2012) was an FMS certified specialist and had four years' experience with the utilisation of the test, so this should not have been a factor in this study. In contrast to the other studies, Teyhen et al. (2012) found only moderate agreement for the ASLR for both inter and intra-rater reliability (kappa 0.69, 0.60 respectively), however percentage of agreement for the ALSR in the Teyhen et al. (2012) study showed excellent agreement with agreement percentage of 84% for inter rater agreement and 80% for intra-rater agreement. When looking at the individual scores in the Teyhen et al. (2012) study only one of the 64 subjects scored by rater one on the first day of testing, was given a score of one for this test and no subjects scored a zero. In this case, it may be that the kappa statistic has under-estimated rater agreement for the ALSR, due to the sensitivity of kappa to prevalence and the lack of spread across the scoring categories.

The FMS includes two measures of flexibility or mobility namely the ASLR and the Shoulder Mobility Test, whereas the NMST has only the one measure of flexibility. The ASLR is a static test that scores on measurement, rather than a movement based test that scores on movement quality. This inherent difference may explain the excellent-perfect agreement seen for the ASLR when compared to other parts of the NMST. Whether a static measure of flexibility is needed within the NMST can be questioned. It may be argued that decreased flexibility or tissue extensibility would limit the ability of an athlete/subject to demonstrate correct movement patterning and that this could therefore be picked up in another NMST test. For example decreased hamstring length that would obstruct the ASLR, would also hinder the ability to get the hips past 90 degrees whilst maintaining a neutral lumbar spine in the bend and pull of the MCS. This would make the addition of the ASLR to the NMST superfluous. It must also be noted that the elimination of the static non movement based components of the NMST, namely the SEBT and the ASLR, may decrease the reliability of the screening tool. Additional research is needed to ascertain if the ASLR is able to predict increased injury risk.

5.7 Limitations

There are a number of relevant limitations in the current study. Firstly whether the screen can be reliably applied to participants of similar sports, such as basketball, or to other age groups cannot be determined by the present study and further research would be required in this area.

Secondly, rater and subject numbers of this study could also be viewed as a limitation. Future studies that utilise higher participant and rater numbers are needed to support the results of the current study and increase the evidence base of the NMST and the components.

Thirdly, the overall NMST score does not include the SEBT scores. This is due to the fact that all other individual tests that make up the NMST are scored out of three on an ordinal scale, whereas the SEBT reach distance is scored on an open ended continuous scale. To be able to add the SEBT score to the NMST overall score poor, fair and normal reach distances, or side to side differences need to be ascertained. As this was beyond the scope of the present study, we decided instead to present NMST overall score and SEBT (right) and SEBT (left) data separately.

Fourthly, the results of the current study are not linked to any injury incidence or exposure data which needs to be done to establish if there is a link between increased injury risk and poor NMST score, as has been done in previous studies looking at the FMS. This was beyond the scope of this thesis and it is recommended that this be done in a future study now the reliability of the NMST has been established in this sporting population.

Fifthly, the higher number of repetitions of the bend and pull and lunge and twist, may have affected the repeatability of the participant’s performance and therefore viewed as a limitation.

Conclusion

The NMST has been shown by this study to have excellent inter and intra-rater agreement for overall NMST score in a cohort of adolescent netball players. The screening tool can be administered reliably by raters with varying levels of clinical experience. However decreased variability of scoring and improvements in MDC score are required to ascertain meaningful clinical change. On-going research needs to be undertaken to establish whether the NMST is a valid tool in identifying increased injury risk for netball players and whether the NMST can be administered reliably by other health and sport medicine professionals.

References

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine, 26*(4), 217-238.

Australia, S. M., &. (2009). *Growing up with exercise: Teenage girls staying active* [Women in Sport Fact Sheet]. Victoria, Australia: Jolimont.

Australian Bureau of Statistics. (2009). *Children's participation in cultural and leisure activities.* Sydney, Australia

Brukner, P. D., & Khan, K. (2012). *Clinical sports medicine* (4th ed.). Sydney, Australia: McGraw Hill.

Butler, R. J., Plisky, P. J., Southers, C., Scoma, C., & Kiesel, K. B. (2010). Biomechanical analysis of the different classifications of the functional movement screen deep squat test. *Sports Biomechanics, 9*(4). doi:10.1080/14763141.2010.539623

Chmielewski, T. L., Hodges, M. J., Horodyski, M.-B., Bishop, M. D., Conrad, B. P., & Tillllman, S. M. (2007). Investigation of clinician agreement in evaluating movement quality during unilateral lower extremity functional tasks: a comparison of 2 rating methods. *Journal of Orthopaedic & Sports Physical Therapy, 37*(3), 122-129.

Cook, G., Burton, L., & Hoogenboom, B. (2006). Pre-participation screening: the use of fundamental movements as an assessment of function – part 2. *North American Journal of Sports Physical Therapy, 1*(3), 132-139.

Davis Hammonds, A. L., Laudner, K. G., McCaw, S., & McLoda, T. A. (2012). Acute lower extremity running kinematics after a hamstring stretch. *Journal of Athletic Training, 47*(1), 5-14.

Demura, S., & Yamada, T. (2010). Proposal for a practical star excursion balance test using three trials and four directions. *Sport, Science and Health, 6*(1), 1-8. doi:10.1007/s11332-010-0089-3

Dennis, R. J., Finch, C. F., Elliot, B. C., & Farhart, P. J. (2008). The reliabiltiy of musculoskeletal screening tests used in cricket. *Physical Therapy in Sport, 9*, 25-33. doi:10.1016/j.ptsp.2007.09.004

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health, 52*(6), 377-384.

Ekegren, C. L., Miller, W. C., Celebrini, R. G., Eng, J. J., & MacIntyre, D. L. (2009). Reliability and validity of observational risk screening in evaluating dynamic knee valgus. *Journal of Orthopaedic & Sports Physical Therapy, 39*(9), 665-674. doi:10.2519/jospt.2009.3004

Feinstein, A. R., & Cicchetti, D. V. (2000). High agreement but low kappa: i. The problems of two paradoxes. *Journal of Clinical Epidemiology, 43*(6), 543-549.

Finch, C. F. (1996). *Evaluating injury prevention strategies*: Proceedings of Australian Conference of Science and Medicine in Sport.

Finnoff, J. T., Peterson, V. J., Hollman, J. H., & Smith, J. (2009). Intrarater and interrater reliability of the balance error scoring system (BESS). *American Academy of Physical Medicine and Rehabilitation, 1*(1), 50-54.

Flood, L., & Harrison, J. E. (2009). Epidemiology of basketball and netball injuries that resulted in hospital admission in Australia, 2000-2004. *The Medical Journal of Australia, 190*(2), 87-90.

Ford, K. R., Myer, G. D., & Hewett, T. E. (2003). Valgus knee motion during landing in high school female and male basketball players. *Medicine and Science in Sports and Exercise, 35*, 1745–1750. doi:10.1249/01.MSS.0000089346.85744.D9

Gabbe, B. J., Bennell, K. L., Wajswelner, H., & Finch, C. F. (2004). Reliability of common lower extremity musculoskeletal screening tests. *Physical Therapy in Sports, 5*(2), 90-97. doi:10.1016/j.ptp.2004.01.003

Gribble, P. A. (2003). The star excursion balance test as a measurement tool. *Athletic Therapy Today, 8*(2), 46-47.

Gribble, P. A., & Hertel, J. (2003). Considerations for normalizing measures of the star excursion balance test. *Measurement in Physical Education and Exercise Science, 7*(2), 89-100.

Gribble, P. A., Hertel, J., & Plisky, P. J. (2012). Using the star excursion balance test to assess dynamic postural-control deficits and outcomes in lower extremity injury: a literature and systematic review. *Journal of Athletic Training, 47*(Supp 1), 339-357.

Gwet, K. (2008). Computing inter-rater reliability: How to estimate the level of agreement between multiple raters. *British Journal of Mathematical and Statisical Epidemiology, 61*(1), 29.

Hartig, D. E., & Henderson, J. M. (1999). Increasing hamstring flexibility decreases lower extremity overuse injuries in military basic trainees. *The American Journal of Sports Medicine, 27*(2), 173-176.

Harty, C. M., DuPont, C. E., Chmielewski, T. L., & Mizner, R. L. (2011). Intertask comparison of frontal plane knee position and moment in female athletes during three distinct movement tasks. *Scandinavian Journal of Medicine & Science in Sports*, *21*(1), 98-105.

Hayen, A., Dennis, R. J., & Finch, C. F. (2007). Determining the intra- and inter-observer relaibility of screening tools used in sports injury research. *Journal of Science and Medicine in Sport, 10,* 201-210. doi:10.1016/j.,sama.2006.09.002

Herrington, L., Hatcher, J., Hatcher, A., & McNicholas, M. (2009). A comparison of star excursion balance test reach distances between ACL deficient patients and asymptomatic controls. *The Knee, 16*, 149-152.

Hertel, J., Braham, R. A., Hale, S. A., & Olmsted-Kramer, L. C. (2006). Simplifying the star-excursion balance test: analyses of subjects with and without chronic ankle instability. *Journal of Orthopaedic & Sports Physical Therapy, 36*, 131-137.

Hertel, J., Miller, S. J., & Denegar, C. R. (2000). Intratester and intertester reliabilty during the star excursion balance tests. *Journal of Sports Rehabilitation, 9*, 104-116.

Hewett, T. E., Myer, G. D., Ford, K. R., Heidt, R. S., Colosimo, A. J., McLean, S. G., ... Succop, P. (2005). Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes. *The American Journal of Sports Medicine, 33*(4), 492-501. doi:10.1177/0363546504269591

Hewett, T. E., Myer, G. D., Ford, K. R., & Slauterbeck, J. R. (2006). Preparticipation physical examination using a box drop vertical jump test in young athletes. The effects of puberty and sex. *Clinical Journal of Sports Medicine*, *16*(4), 298-304.

Hopper, D., & Elliot, B. (1993). Lower limb and back injury patterns of elite netball players. *Sports Medicine, 16*(2), 148-162.

Hopper, D., Elliot, B., & Lalor, J. (1995). A descriptive epidemiology of netball injuries during competition: a five year study. *British Journal of Sports Medicine, 29*(4), 223-228.

Hunt, T. N., Ferrara, M. S., Bornstein, R. A., & Baumgartner, T. A. (2009). The reliability of the modified balance error scoring system. *Clinical Journal of Sports Medicine*, *19*(6), 471-475.

Johnston, M. (2013). Jump in netball ACC injuries. *New Zealand Herald*. Retrieved from http://www.nzherald.co.nz/netball/news/article.cfm?c_id=69&objectid=10878635

Karim, A., Millet, V., Olson, S., & Morganthaler, A. (2011). Inter-rater reliability of a musculoskeletal screen as administered to female professional contemporary dancers. *Work, 40*(3), 281-288. doi:10.3233/WOR-2011-1231

Kibler, W. B., Chandler, J., Uhl, T., & Maddux, R. E. (1989). A musculoskeletal approach to the preparticipation physical examination. Preventing injury and improving performance. *The American Journal of Sports Medicine, 17*(4), 525-531. doi:10.1177/036354658901700413

Kiesel, K. B., & Honarbakhsh, B. (2009). *The functional movement screen* TM *workshop*. Vancouver, Canada: Sport Physiotherapy Canada.

Kiesel, K. B., Plisky, P. J., & Butler, R. J. (2011). Functional movement test scores improve following a standardized off-season intervention program in professional football players. *Scandinavian Journal of Medicine & Science in Sports, 21*, 287-292. doi:10.1111/j.1600-0838.2009.01038.x

Kiesel, K. B., Plisky, P. J., & Voight, M. L. (2007). Can serious injury in professional football be predicted by a preseason functional movement screen? *North American Journal of Sports Physical Therapy, 2*(3), 147-158.

Kinzey, S. J., & Armstrong, C. W. (1998). The reliability of the star-excursion test in assessing dynamic balance. *Journal of Orthopaedic & Sports Physical Therapy, 27*(5), 356-360.

Kritz, M., Cronin, J., & Hume, P. (Forthcoming: 2014). Reliability of the movement competency screen (MCS). *Australian Journal of Strength and Conditioning. Retrieved from: Personal communication*. doi:10.1016/jsams.2009.10.104

McGuine, T. A., Greene, J. J., Best, T., & Leverson, G. (2000). Balance as a predictor of ankle injuries in high school basketball players. *Clinical Journal of Sports Medicine, 10*, 239-244.

McHugh, M. P., Tyler, T. F., Tetro, D. T., Mullaney, M. J., & Nicholas, S. J. (2006). Risk factors for noncontact ankle sprains in high school athletes : The role of hip strength and balance ability. *The American Journal of Sports Medicine, 34*(3), 464-470. doi:10.1177/0363546505280427

McManus, A., Stevenson, M. R., & Finch, C. F. (2006). Incidence and risk factors for injury in non-elite netball. *Journal of Science and Medicine in Sport, 9*, 119-124.

Meeuwisse, W. (1994). Assessing causation in sports injury: a multifactorial model. *Clinical Journal of Sports Medicine, 4*(3), 166-170.

Miller, A., & Callister, R. (2009). Reliable lower limb musculoskeletal profiling using easily operated, portable equipment. *Physical Therapy in Sport, 10,* 30-37. doi:10.1016/j.ptsp.2008.10.003

Milner, C. E., Westlake, C. G., & Tate, J. J. (2011). Test-retest reliability of knee biomechanics during stop jump landings. *Journal of Biomechanics, 44*, 1814-1816. doi:10.1016/j.jbiomech.2011.04.005

Minick, K. I., Kiesel, K. B., Burton, L., Taylor, A., Plisky, P. J., & Butler, R. J. (2010). Interrater realibility of the functional movment screen. *Journal of Strength & Conditioning Research, 24*(2), 479-486.

Munro, A. G., & Herrington, L. (2010). Between-session reliability of the star excursion balance test. *Physical Therapy in Sport*, *11*, 128-132. doi:10.1016/j.ptsp.2010.07.002

Munro, A. G., Herrington, L., & Carolan, M. (2012). Reliability of 2-dimensional video assessment of frontal-plane dynamic knee valgus during common athletic screening tasks. *Journal of Sport Rehabilitation, 21*(1), 7-11.

Myklebust, G., & Bahr, R. (2005). Return to play guidelines after anterior cruciate ligament surgery. *British Journal of Sports Medicine, 39*, 127-131. doi:10.1136/bjsm.2004.010900

Myklebust, G., Engebretsen, L., Brækken, I. F., Skjølberg, A., Olsen, O.-E., & Bahr, R. (2003). Prevention of anterior cruciate ligament injuries in female team handball players: a prospective intervention study over three seasons. *Clinical Journal of Sports Medicine, 13*(2), 71-78.

Nagano, Y., Fukano, M., Itagaki, K., Li, S., Miyakawa, S., & Fukubayashi, T. (2010). Influence of lower limb clinical physical measurements of female athletes on knee motion during continuous jump testing. *The Open Sports Medicine Journal 4*, 134-139.

Netball New Zealand. (2013). *Netball New Zealand. The official site*. Retrieved March 31, 2013, from http://www.mynetball.co.nz/index.php/community-netball/59/110

O'Sullivan, K., McAuliffe, S., & DeBurca, N. (2012). The effects of eccentric training on lower limb flexibility: a systematic review. *British Journal of Sports Medicine, 46*, 838-845. doi:10.1136/bjsports-2011-090835

Olmsted, L. C., Carcia, C. R., Hertel, J., & Shultz, S. J. (2002). Efficacy of the star excursion balance tests in detecting reach deficits in subjects with chronic ankle instability. *Journal of Athletic Training, 37*(4), 501-506.

Onate, J. A., Cortes, N., Welch, C., & van Lunen, B. L. (2010). Expert versus novice interrater reliability and criterion validity of the landing error scoring system. *Journal of Sport Rehabilitation, 19*(1), 41-56.

Onate, J. A., Dewey, T., Kollock, R. O., Thomas, K. S., Van Lunen, B. L., DeMaio, M., & Ringleb, S. I. (2012). Real-time intersession and interrater reliabilty of the functional movement screen. *Journal of Strength & Conditioning Research, 26*(2), 408-415.

Ostenberg, A., & Roos, H. (2000). Injury risk factors in female European football. A prospective study of 123 players during one season. *Scandinavian Journal of Medicine & Science in Sports, 10*, 279-285.

Otago, L., & Peake, J. (2007). The role of insurance data in setting priorities for netball injury prevention strategies. *Journal of Science and Medicine in Sport, 10*(2), 105-109.

Padua, D. A., Boling, M. C., DiStephano, L. J., Onate, J. A., Beulter, A. I., & Marshall, S. W. (2011). Reliabilty of the landing error scoring system-real time, a clinical assessment tool of jump-landing biomechanics. *Journal of Sport Rehabilitation, 20*, 145-156.

Padua, D. A., Marshall, S. W., Boling, M. C., Thigpen, C. A., Garret, W. E., & Beutler, A. I. (2009). The landing error scoring system (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: the JUMP-ACL study. *The American Journal of Sports Medicine, 37*(10), 1996-2002.

Plisky, P. J., Gorman, P. P., Butler, R. J., Kiesel, K. B., Underwood, F. B., & Elkins, B. (2009). The reliability of an instrumented device for measuring components of the star excursion balance test. *North American Journal of Sports Physical Therapy, 4*(2), 92-99.

Plisky, P. J., Rauh, M. J., Kaminski, T. W., & Underwood, F. B. (2006). Star excursion balance test as a predictor of lower extremity injury in high school basketball players. *Journal of Orthopaedic & Sports Physical Therapy, 36*(12), 911-919.

Riffenburgh, R. H. (2012). *Statistics in Medicine* (3rd ed.)

Robinson, R., & Gribble, P. A. (2008a). Kinematic predictors of performance on the star excursion balance test. *Journal of Sport Rehabilitation, 17*, 347-357.

Robinson, R., & Gribble, P. A. (2008b). Support for a reduction in the number of trials needed for the star excursion balance test. *Archives of Physical and Medical Rehabilitation, 89*, 364-370. doi:10.1016/j.apmr.2007.08.139

Ross, M. D., Langford, B., & Whelan, P. J. (2002). Test-retest reliability of 4 single-leg horizontal hop tests. *Journal of Strength & Conditioning Research (Allen Press Publishing Services Inc.), 16*(4), 617-622.

Rumsey, D. (2011). *Statistics for Dummies*

Sale, B. (1991). *Junior sport in South Australia*: Commissioned by the Junior Sports Unit of the South Australian Sports Institute.

Samples, P. (1986). Pre-participation exams: are they worth the time and trouble? *Physician Sportsmedicine, 14*(1), 180-187.

Saunders, B., & Eggart, J. S. (Eds.). (1979). *Guidelines for preseason athletic participation evaluations*. Alexandria, VA: American Physical Therapy Association.

Schneiders, A. G., Davidsson, A., Horman, E., & Sullivan, S. J. (2011). Functional movement screen normative values in a young active population. *The International Journal for Sports Physical Therapy, 6*(2), 75-82.

Schwenk, M., Gogulla, S., Englert, S., Czempik, A., & Hauer, K. (2012). Test–retest reliability and minimal detectable change of repeated sit-to-stand analysis using one body fixed sensor in geriatric patients. *Physiological Measurement, 33*, 1931-1946. doi:10.1088/0967-3334/33/11/1931

Sheehan, D. P., Lafave, M. R., & Katz, L. (2011). Intra-rater and inter-rater reliability of the balance error scoring system in pre-adolescent school children. *Measurement in Physical Education and Exercise Science, 15*(3), 234-243. doi:10.1080/1091367X.2011.568368

Shultz, R., Mooney, K., Anderson, S., Marcello, B., Garza, D., Matheson, G. O., & Besier, T. (2011). Functional movement screen: Inter-rater and subject reliability. *British Journal of Sports Medicine, 45*, 374. doi:10.1136/bjsm.2011.084038

Shultz, S. J., Nguyen, A.-D., Windley, T. C., Kulas, A. S., Botic, T. L., & Beynnon, B. D. (2006). Intratester and intertester reliability of clinical measures of lower extremity anatomic characteristics: implications for multicenter studies. *Clinical Journal of Sports Medicine, 16*(2), 155-161.

Smith, H. C., Johnson, R. J., Shultz, S. J., Tourville, T., Holterman, L.-A., Slauterbeck, J., ... Beynnon, B. D. (2012). A prospective evaluation of the landing error scoring system (LESS) as a screening tool for anterior cruciate ligament injury risk. *American Journal of Sports Medicine*, *40*(3), 521-526.

SPARC. (2001). *Adult participation in sport and active leisure*. Retrieved 28/03, 2013, from http://www.activenzsurvey.org.nz/Results/NZ-Sport-and-Physical-Surveys-9701/SPARC-Facts-97-01/Part-3/Adult-participation-in-sport-and-active-leisure/

SPARC. (2009). *Sport and recreation profile: Netball – findings from the 2007/08 active NZ survey*. Wellington: SPARC.

Steele, J. R. (1990). Biomechanical factors affecting performance in netball. Implications for improving performance and injury reduction. *Sports Medicine*, *10*(2), 88-102.

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), 530-540. doi:10.2519/jospt.2012.3838

Trojian, T. H., & McKeag, D. B. (2006). Single leg balance test to identify risk of ankle sprains. *British Journal of Sports Medicine*, *40*, 610-613. doi:10.1136/bjsm.2005.024356

Urdan, T. C. (2010). *Statistics in Plain English* (3rd ed.). New York, NY: Taylor and Francis Group.

Wang, H.-K., Chen, C.-H., Shiang, T.-Z., Jan, M.-H., & Lin, K.-H. (2006). Risk-factor analysis of high school basketball–player ankle injuries: a prospective controlled cohort study evaluating postural sway, ankle strength, and flexibility. *Archives of Physical and Medical Rehabilitation*, *87*. doi:10.1016/j.apmr.2006.02.024

Weldon, S. M., & Hill, R. H. (2003). The efficacy of stretching for prevention of exercise-related injury: a systematic review of the literature. *Manual Therapy*, *8*(3), 141-150. doi:10.1016/S1356-689X(03)00010-9

Whatman, C., Hing, W., & Hume, P. (2012). Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. *Physical Therapy in Sport*, *13*(2), 87-96. doi:10.1016/j.ptsp.2011.07.001

WHO Consultation on Obesity. (1999). *Obesity: preventing and managing the growing epidemic: report of a WHO consultation*. Geneva, Switzerland.

Appendix A: Ethical Approval

MEMORANDUM

Auckland University of Technology Ethics Committee (AUTEC)

To:	Duncan Reid
From:	**Dr Rosemary Godbold** Executive Secretary, AUTEC
Date:	18 November 2011
Subject:	Ethics Application Number 11/296 **Does a netball screening tool predict injury in secondary school netball players?**

Dear Duncan

Thank you for providing written evidence as requested. I am pleased to advise that it satisfies the points raised by the Auckland University of Technology Ethics Committee (AUTEC) at their meeting on 31 October 2011 and I have approved your ethics application. This delegated approval is made in accordance with section 5.3.2.3 of AUTEC's *Applying for Ethics Approval: Guidelines and Procedures* and is subject to endorsement at AUTEC's meeting on 12 December 2011.

Your ethics application is approved for a period of three years until 17 November 2014.

I advise that as part of the ethics approval process, you are required to submit the following to AUTEC:

- A brief annual progress report using form EA2, which is available online through http://www.aut.ac.nz/research/research-ethics/ethics. When necessary this form may also be used to request an extension of the approval at least one month prior to its expiry on 17 November 2014;
- A brief report on the status of the project using form EA3, which is available online through http://www.aut.ac.nz/research/research-ethics/ethics. This report is to be submitted either when the approval expires on 17 November 2014 or on completion of the project, whichever comes sooner;

It is a condition of approval that AUTEC is notified of any adverse events or if the research does not commence. AUTEC approval needs to be sought for any alteration to the research, including any alteration of or addition to any documents that are provided to participants. You are reminded that, as applicant, you are responsible for ensuring that research undertaken under this approval occurs within the parameters outlined in the approved application.

Please note that AUTEC grants ethical approval only. If you require management approval from an institution or organisation for your research, then you will need to make the arrangements necessary to obtain this.

When communicating with us about this application, we ask that you use the application number and study title to enable us to provide you with prompt service. Should you have any further enquiries regarding this matter, you are welcome to contact me by email at ethics@aut.ac.nz or by telephone on 921 9999 at extension 6902.

On behalf of AUTEC and myself, I wish you success with your research and look forward to reading about it in your reports.

Yours sincerely

Dr Rosemary Godbold
Executive Secretary
Auckland University of Technology Ethics Committee

Cc: Rebecca Vanweerd becsvw@hotmail.com

Appendix B: Studies excluded from literature review and reason for exclusion

Author and publication date	Article Title	Reason for Exclusion
Dennis, Finch, Elliot, & Farhart. (2008)	The reliability of musculoskeletal tests used in cricket.	Static tests only.
Finnoff, Peterson, Hollman, & Smith. (2009)	Intra-rater and inter-rater reliability of the Balance Error Scoring System (BESS).	Screen developed and used for concussion assessment rather than injury prevention.
Gabbe, Bennell, Wajswelner, & Finch. (2004)	Reliability of common lower extremity musculoskeletal screening tests.	Static tests only. Not movement based.
Harty, DuPont, Chmielewski, & Mizner. (2011)	Inter-task comparison of frontal plane knee position and moment in female athletes during three distinct movement tasks.	Looking at relationship between jumping tasks rather than rater reliability.
Hunt, Ferrara, Bornstein, & Baumgartner. (2009)	The reliability of the Modified Balance Error Scoring System.	Screen developed and used for concussion assessment rather than injury prevention.
Milner, Westlake & Tate. (2011)	Test-retest reliability of knee biomechanics during stop jump landings	Looking at reliability of performance rather than rater reliability
Munro & Herrington. (2010)	Between-session reliability of the star excursion balance test.	Looking to ascertain number of trials needed before performance is consistent and learning effect is negated rather than rater reliability.
Munro, Herrington & Carolan. (2012)	Reliability of 2-dimensional video assessment of frontal-plane dynamic knee valgus during common athletic screening tasks.	Not looking at rater reliability.
Ross. Langford & Whelan. (2002)	Test-retest reliability of 4 single-leg horizontal hop tests.	Not looking at rater reliability.

Shultz et al. (2006)	Lower extremity anatomical characteristics.	Static tests only.
Shultz et al. (2011)	Functional Movement Screen: inter-rater and subject reliability.	Poster session only no full text article.
Sheehan, Lafave & Katz. (2011)	Intra-rater and Inter-rater reliability of the Balance Error Scoring System in pre-adolescent school children.	Screen developed and used for concussion assessment rather than injury prevention.

Appendix C: Modified Downs and Black for Reliability Studies

1) Is the hypothesis/aim/objective/main outcomes of the study clearly described?

If the main outcomes are first mentioned in the results section then the answer is no

Yes	1
No	0

2) Are the characteristics of patients included in the study clearly described?

This includes inclusion and exclusion criteria for cohort studies and trials.

Yes	1
No	0

3) Is the screening tool of interest clearly described?

Screening tools should be clearly described.

Yes	1
No	0

4) Are the distributions of principle confounders clearly described?

Conditions in which subjects are tested should be consistent between subjects. If not these need to be listed.

Yes	2
Partially	1
No	0

5) Are the main findings of the study clearly described?

Simple outcome data (including denominators and numerators) should be reported for all major findings so that the reader can check the major analyses and conclusions. This question does not cover statistical tests or analyses.

Yes	1
No	0

6) Have all important adverse events that may be a consequence of the intervention been reported?

This should be answered yes if the study demonstrated that there was a comprehensive attempt to measure adverse events.

Yes	1

No	0

7) Have the characteristics of patients lost to re-testing been described? **(Intra-rater only)**

This should be answered yes where there were no losses to re-testing or where losses to re-testing were so small that findings would be unaffected by their inclusion.

Yes	1
No	0

8) Have actual statistical values such as kappa or ICC been reported (e.g. 0.035 rather than poor)

Yes	1
No	0

9) Were the subjects asked to (and prepared to) participate in the study representative of the entire population from which they were recruited?

The study must identify the source population for patients and describe how patients were selected. Patients would be representative if they comprised the entire source population, an unselected sample of consecutive patients or a random sample. Random sampling is only feasible where a list of all members of the relevant population exists. Where a study does not report the proportion of the source population from which the patients are derived, the question should be answered as unable to determine. Where a convenience sample has been undertaken answer unable to determine.

Yes	1
No	0
Unable to determine	0

10) Were assessors adequately trained in standardised testing procedures and equipment?

Yes	1
No	0
Unable to determine	0

11) Did assessors have a similar background in the implementation and scoring of the screening test? **(Inter-rater only)**

Yes	1
No	0
Unable to determine	0

Internal Validity Bias

12) Was an attempt made to blind raters to either their initial scores, or the scores of the other rater?

Yes	1
No	0
Unable to determine	0

13) If any results of the study were based on data dredging was this made clear?

Any analyses that had not been planned at the outset of the study should be clearly indicated. If no retrospective unplanned subgroup analyses were reported then answer yes.

Yes	1
No	0
Unable to determine	0

14) Was adequate time and practice to familiarise the physical tests, or detailed instructions given to subjects?

Yes	1
No	0
Unable to determine	0

15) Was the screen implemented using standardised procedures?

Tests were not randomised in order which may affect reliability of results. If no standardised procedures exist answer yes

Yes	
No	
Unable to determine	

16) Were the statistical tests used to assess the main outcomes appropriate?

The statistical techniques used must be appropriate to the data. ICC and kappa values used for reliability.

Yes	
No	
Unable to determine	

17) Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?

Were physical attributes of individual subjects that may affect the result of the screen controlled for? If this was not applicable answer yes.

Yes	
No	
Unable to determine	

Power

18) Was the sample size appropriate to measure reliability?

Did studies justify why their choice of sample size was appropriate. If no mention of why or how a sample size was chosen the answer is no

Yes	1
No	0

Appendix D: Netball Movement Screening Tool

Movement Competency Screen

Pattern	Primary	Secondary	Score	Comments
Squat	O Shoulders O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth O Balance	1 2 3	
Lunge (left)	O Balance O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth	1 2 3	
Lunge (right)	O Balance O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth		
Twist (left)	O Shoulders O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth O Balance	1 2 3	
Twist (right)	O Shoulders O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth O Balance		
Bend	O Shoulders O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth O Balance	1 2 3	
Pull	O Shoulders O Lumbar O Hips O Depth	O Head O Knees O Ankles / Feet O Balance	1 2 3	
Push Up	O Head O Shoulders O Lumbar O Depth	O Hips O Knees O Ankles / Feet O Balance	1 2 3	
Single Leg Squat (left)	O Depth O Lumbar O Hips O Ankles / Feet	O Head O Shoulders O Knees O Balance	1 2 3	
Single Leg Squat (right)	O Shoulders O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth O Balance		

Jump Patterns

Pattern	Primary	Secondary	Score	Comments
Vertical Jump A	O Shoulders O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth O Balance	1 2 3	
Vertical Jump B (left)	O Shoulders O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth O Balance	1 2 3	
Vertical Jump B (right)	O Shoulders O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth O Balance		
Broad Jump (left)	O Shoulders O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth O Balance	1 2 3	
Broad Jump (right)	O Shoulders O Lumbar O Hips O Ankles / Feet	O Head O Knees O Depth O Balance		

Considerations: The numbers in the PRIMARY and SECONDARY columns depict the number of areas that were marked during the screen. Select the 1, 2 or 3 in the Score columns after adding up the checked areas for each pattern. For the vertical jump B and broad jump, the lower of the two scores is recorded.

Scoring Instructions		
Score	**Primary**	**Secondary**
1	**2+ and / or 4**	
2	**1 and / or 0-3**	
3	**0 and / or 0**	

Star Excursion Balance Test

Scoring	Left	Right
Measurement A		
Measurement PL		
Measurement PM		
Leg length		
Score		

Score is equated by the following: $\frac{[(A + PL + PM)]}{(LLx3)} \times 100$

Active straight leg raise

Scoring	
Left	1 2 3
Right	1 2 3

Scoring Summary

Screening Test	Score
MCS	
Jumping Patterns	
SEBT	
ASLR	
Total Score:	

Nb: The SEBT score is not included in the overall NMST score due to the continuous nature of it’s scoring compared to the ordinal ranking used in the other tests. Further research needs to be undertaken to ascertain a way to rank the SEBT in an ordinal fashion.

Appendix E: Netball Movement Screening Tool Instructional Sheet.

Movement Competency Screen

Squat

Feet hip distance apart, hands behind head. Squat down as deep as possible and return to the starting position.

Lunge and twist

Arms crossed over the body. Lunges the back knee down and then twist towards the forward leg. Return to starting position.

Bend and Pull

Feet hip distance apart, very small bend in the knees. Bend forward through the hips then pull elbows back in a rowing motion retracting the scapula. .

Push up

Start with hands under shoulders and up on toes. Lower chest as close to the floor as possible and then push up to return to starting position.

Single leg squat

Hands behind head, standing on one leg, perform a single leg squat as deep as able and return to starting position.

Jumping components

Vertical Jump A

Hands on hips, jump up vertically as high as possible and land on both feet.

Vertical Jump B

Hands on hips, jump up vertically as high as possible and land on one foot only. Perform on left and right sides.

Broad jump

Standing feet hip width, jump forward off both legs as far as possible and lands on one leg only. Perform on left and right sides

Star Excursion Balance Test

Four pieces of tape are laid down on the ground 45 degrees apart from each other, forming eight rays. The eight rays are named for the reach direction they represent based on their anatomical alignment from the midpoint (anterior, anteromedial, anterolateral, lateral, medial, posterior, posterolateral and posteromedial). For the purposes of the NMST test, only the anterior (A), posteromedial (PM) and posterolateral (PL) reach directions utilised. The midpoint is marked where the four pieces of tape intersect and is where the test foot must remain throughout testing.

The arch of the test foot is placed on the midpoint of the star and hands are on hips. The subject is instructed to reach the opposite foot along the chosen reach direction as far as possible and lightly touch it down. The point on the line where the athletes touches down is marked. Each direction is repeated three times and the furthest reach direction is recorded. If reach foot is used for considerable support at any time, support foot is moved, hands come off the hips or subject loses balance, the trial is disregarded and repeated.

Leg length (LL) is measured from distal aspect of anterior superior iliac spine to the distal end of the lateral malleolus. SEBT score is calculated using the following score calculation:

$$\frac{[(A + PL + PM)]}{(LLx3)} \times 100$$

Active Straight Leg Raise

Starting position is supine on the floor. The test leg is actively straightened then lifted up into the air as far as possible keeping the test leg straight and the non-test leg in contact with the ground throughout the entire testing procedure. Hips are to remain in neutral rotation throughout testing. Prior to testing two lines are marked on the subject's leg to divide the leg into three zones for scoring. One line is drawn at mid patella and one at mid-thigh (half way between the ASIS and mid patella). Zone 1 was demarcated as distal to mid patella, Zone 2 between mid-patella and mid-thigh and Zone 3 superior to mid-thigh. The ASLR is scored by aligning a ski pole vertically with the medial malleolus of the test leg and recording which zone of the non-test leg the ski pole was situated in. A three was scored if the pole was situated in Zone 3, a two was scored if the pole was situated in Zone 2 and a one was scored if the pole was situated in Zone 1.

Appendix F: Scoring Criteria for NMST

Pattern	Head	Shoulder/Thoracic Spine	Lumbar Spine	Hips	Knees	Ankles	Feet	Balance	Depth
Squat	Held in neutral position, appears centrally aligned	Thoracic extension evident, shoulders kept down and away from ears. Elbows are in-line with ears during movement	Neutral curve position	Horizontally aligned and mobile. Move back and down during flexion	Aligned with hips and feet during flexion	Mobility allows adequate dorsiflexion during hip and knee flexion	Stable with heels grounded during lower limb flexion	Evenly distributed	Tops of thighs appear parallel with floor
Lunge and Twist	Held in neutral position, appears centrally aligned	Held down and away from ears. Majority of rotation appears to occur through the thoracic spine	Held in a neutral curve position. Rotation and/or lateral flexion does not occur during trunk twisting	Mobile and stable to prohibit elevation and depression during rotation	Aligned with hips and feet during flexion and do not move laterally with rotation	Mobility allows adequate dorsiflexion during hip and knee flexion	Heel of lead leg in contact with the floor, trail foot flexed and balanced on forefoot	Maintained for each leg	Lead thigh parallel with the ground
Bend and Pull	Held in neutral position, appears centrally aligned	Held down and away from ears. Scapulae move balanced and rhythmic. During arm flexion scapulae protract. Scapulae are not excessively abducted during arm extension	Held in a neutral curve position throughout trunk flexion	Facilitate trunk flexion	Extended	N/A	Pointing straight	Maintained	75-90 degrees of trunk flexion acheived

Pattern	Head	Shoulder/Thoracic Spine	Lumbar Spine	Hips	Knees	Ankles	Feet	Balance	Depth
Push Up	Held in neutral position, appears centrally aligned	Held down and away from ears. Scapulae move balanced and rhythmic and are not excessively abducted during arm extension	Held in a neutral curve position	Held in line with the body during arm flexion and extension	Extended	N/A	Straight, not falling out or in	N/A	Chest touches floor
Single leg squat	Held in neutral position, appears centrally aligned	Held down and away from ears. Elbows in line with ears	Held in neutral curve position	Mobile and facilitate flexion and stable to minimise weight shift over stance leg	Aligned with hips and feet during flexion	Mobility allows adequate dorsiflexion during hip and knee flexion	Stable with heels grounded during lower limb flexion	Maintained on each leg	Top of thighs appear parallel with floor
Vertical Jump A	Held in neutral position, appears centrally aligned	Shoulder held down and away from ears, thoracic spine stable	Held in neutral curve no hyper-extension	Neutral rotation flexed on landing	Aligned over second toe. Flexion greater than 30 degrees	Aligned with knee and hip	Aligned, lands on toes	Remains balanced during take-off and landing. Holds landing for two seconds	Hips and knees are flexed 30-45 degrees at landing
Vertical Jump B	Held in neutral position, appears centrally	Shoulder held down and away from ears, thoracic spine stable	Held in neutral curve no hyper-extension	Neutral rotation flexed on landing	Aligned over second toe. Flexion greater than 30 degrees	Aligned with knee and hip	Aligned, lands on toes	Remains balanced during take-off and landing. Holds landing	Hips and knees are flexed 30-45 degrees at landing

Pattern	Head	Shoulder/Thoracic Spine	Lumbar Spine	Hips	Knees	Ankles	Feet	Balance	Depth
	aligned							for two seconds	
Broad Jump	Held in neutral position, appears centrally aligned	Shoulder held down and away from ears, thoracic spine stable	Held in neutral curve no hyper-extension	Neutral rotation flexed on landing	Aligned over second toe. Flexion greater than 30 degrees	Aligned with knee and hip	Aligned, lands on toes	Remains balanced during take-off and landing. Holds landing for two seconds	Hips and knees are flexed 30-45 degrees at landing