Physiotherapy visual assessment of dynamic alignment during lower extremity functional screening tests

A thesis submitted to Auckland University of Technology

in fulfilment of the requirements for the degree of

Doctor of Philosophy

February 2012

Auckland University of Technology

Faculty of Health and Environmental Sciences

by

Chris Whatman DipPhys MAppSc

.

Primary supervisor: Professor Wayne A. Hing

Secondary supervisor: Professor Patria A. Hume

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

9

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 (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Chris Whatman

February 2012
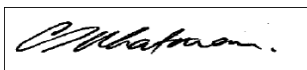
CANDIDATE CONTRIBUTIONS TO CO-AUTHORED WORKS

20th December 2011

All co-authors on the chapters/papers indicated in the following table have approved these for inclusion in Chris Whatman's doctoral thesis.

Chapter publication reference	Author %
CHAPTER 3: Whatman, C., Hing, W., & Hume, P. A. (2011). Kinematics during lower extremity functional screening tests - Are they reliable and related to jogging? *Physical Therapy in Sport, 12*, 22-29.	CW = 85%; WH = 5%; PH = 10%.
CHAPTER 4: Whatman, C., Hing, W., & Hume, P. A. (2011). Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. *Physical Therapy in Sport, in press*. doi:10.1016/j.ptsp.2011.07.001	CW = 85%; WH = 5%; PH = 10%.
CHAPTER 5: Whatman, C., Hume, P. A., & Hing, W. (2012 in review). Kinematics during lower extremity functional screening tests in young athletes – are they reliable and valid? *Physical Therapy in Sport*.	CW = 85%; PH = 10%; WH = 5%.
CHAPTER 6: Whatman, C., Hume, P. A., & Hing, W. (2012 in review). The reliability and validity of physiotherapist visual rating of dynamic pelvis and knee alignment in young athletes. *Physical Therapy in Sport*.	CW = 85%; PH = 10%; WH = 5%.
CHAPTER 7: Whatman, C., Hume, P. A., & Hing, W. (2012 in review). The associations between 2D alignment and 3D kinematics during lower extremity functional tests in young athletes. *Sport Biomechanics*.	CW = 85%; PH = 10%; WH = 5%.
CHAPTER 8: Whatman, C., Hume, P. A., & Hing, W. (2012 in progress). Visual and kinematic assessments of dynamic alignment during lower extremity functional screening tests: A review of literature. To be submitted to *Sports Medicine*.	CW = 85%; PH = 10%; WH = 5%.

APPENDIX 1: Whatman, C., Hing, W., & Hume, P. A. (2008, November). Physiotherapist agreement in rating movement quality during lower extremity functional tests *Sports Medicine New Zealand.* Symposium conducted at the meeting of the New Zealand Conference of Medicine and Science in Sport, Dunedin.	CW = 85%; PH = 10%; WH = 5%.
APPENDIX 2: Whatman, C., Hume, P. A., & Hing, W. (2009, November). Lower extremity functional tests – are they reliable and do they predict actual function? Symposium conducted at the meeting of the New Zealand Conference of Medicine and Science in Sport, Rotorua.	CW = 85%; PH = 10%; WH = 5%.

Mr. Christopher Whatman (CW)   Prof. Wayne Hing (WH)   Prof. Patria Hume (PH)

ACKNOWLEDGEMENTS

I would like to thank the following people for the contribution they have made, and support they have given, to my PhD thesis.

I would like to thank my wife Stella for doing all the hard work during the time of this PhD! – you have been the number one parent and your success is obvious in our two boys who are the nicest young fellas I know! I would also like to thank Reuben and James for their laughter and 'snuggles' which helped keep the PhD in perspective.

I would like to thank my supervisors Professor Wayne Hing and Professor Patria Hume. Wayne your guidance throughout the PhD and your wealth of physiotherapy experience and expertise was of great value. You always made yourself available when needed and gave much appreciated encouragement and advice. We both saw the value in hard work but also shared many laughs which I always appreciated. Patria your work ethic and dedication to her students is astonishing, even when not in the best of health. Your attention to detail and willingness to give of your time is without peer in my experience and was truely appreciated. Your research writing retreats are an annual highlight with so many intangible benefits over and above the advancement of research.

DEDICATION

I wish to dedicate this thesis to my mum. You always set high standards for your boys and encouraged us to complete any task to the best of our ability. Perhaps most importantly you took an interest in how we did and were always available to provide help. Now that I'm a parent I can see that great parents are truely the most valuable resource on the planet!

ETHICAL APPROVAL

Ethical approval for this thesis research was granted by the Auckland University of Technology Ethics Committee (AUTEC). The AUTEC references were:

- Ethics Application Number 08/258 **Relationship of hip strength to running mechanics in young athletes.**

- Ethics Application Number 06/143 **Reliability and validity of physiotherapy screening of lower limb function.**

ABSTRACT

Physiotherapist visual assessment of lower extremity dynamic alignment during functional screening tests is common when assessing clients for risk of injury and during rehabilitation. However the reliability and validity of visual assessment for several functional tests has not been reported, and for other tests needs clarification. The aims of this thesis were to investigate: 1) reliability of kinematics during lower extremity functional screening tests and their association with function (running and landing); and 2) reliability and validity of visual assessment of dynamic alignment during functional tests. Three-dimensional (3D) and/or two-dimensional (2D) kinematics were measured in healthy adults and young athletes during small knee bend (SKB), lunge, hop lunge, step-down and drop jump functional tests. Within-day (ICC ≥0.85) and between-days (ICC ≥0.60) reliability was acceptable for the majority of kinematics. Associations between functional test kinematics and kinematics during running and landing were moderate to very large. Physiotherapists with a range of experience visually rated dynamic alignment during functional tests using segmental and overall body approaches and dichotomous and ordinal scales. Mean intra-rater agreement was moderate to good [Agreement Coefficient 1 (AC1): 0.39 to 0.80] and inter-rater agreement fair to good (AC1: 0.22 to 0.71). Clinical experience, the use of a dichotomous scale and rating the knee position relative to the foot all improved agreement. Agreement on an overall rating was similar to segmental ratings. In young athletes, sensitivity (≥80%) and specificity (≥50%) were acceptable for visual ratings of SKB (double and single leg) but not drop jumps when compared to expert consensus ratings (aided by video slow motion). Experience and slower test velocity improved rating accuracy when rating young athletes [Diagnostic Odds Ratio (DOR): 1.6 to 4.9 times better]. Expert consensus ratings differentiated young athletes with different 2D kinematics (very likely to almost certainly) and 3D hip kinematics (likely to very likely) but not 3D knee abduction. The association between 2D and 3D kinematics during SKB and drop jumps in young athletes ranged from small to very large. Kinematics during lower extremity functional screening tests show sufficient reliability and link to function. Physiotherapist visual rating of dynamic alignment during these tests provides reliable and valid information that should assist in the clinical decision making process.

CHAPTER1: INTRODUCTION AND RATIONALISATION

Background

Assessment of movement quality (neuromuscular control), evaluating multiple joints in a chain rather than assessment of individual joints alone, is an important part of musculoskeletal assessment. From a physiotherapy perspective, assessment of relative strength and flexibility between joints is just as important as assessment of any joint individually. Mottram and Comerford (2008) noted that the traditional assessment of isolated joints and muscles alone is no longer considered adequate and the additional use of functional tests to assess multiple muscles, across multiple joints, in functionally orientated tasks is now common. Sahrmann (2002) has been a proponent of this approach for many years referring to the term movement quality, emphasising that assessment and treatment of musculoskeletal disorders should emphasise quality rather than quantity. Sahrmann suggested poor movement quality (faulty patterns and/or poor control) is the key risk factor for the development of injury as it leads to increased stress on the components of the musculoskeletal system resulting in cumulative microtrauma and musculoskeletal pain syndromes. Sahrmann (2011) recently proposed that evaluation of the movement system should be as regular as dental assessments, reinforcing that quality of movement is the key with precise exercise and modification of functional activities needed to prevent early onset and severity of musculoskeletal pain.

In clinical practice the assessment of movement quality/dynamic alignment is predominantly via visual observation, as physiotherapists do not have access to the equipment or time required for complex biomechanical analysis. The benefit of visual assessment is the ability to immediately evaluate changes in performance and provide feedback. Visual assessment of movement is a key skill required of physiotherapists and has been recognised as an important component of a therapists' decision making process (Bernhardt, Bate, & Matyas, 1998). It provides valuable information that can guide further assessment and is used across a range of clinical areas including the musculoskeletal context (Passier, Nasciemento, Gesch, & Haines, 2010). However the reliability and validity of this common and important assessment technique remains unclear in many situations. The visual rating of upper limb movement and gait in participants with neurological impairments has received most attention. Most studies show that intra-observer reliability is moderate to good and inter-observer reliability is

poor to good (Bernhardt, et al., 1998; Lord, Halligan, & Wade, 1998). However, the reliability and validity of visual rating of lower extremity movement quality in participants with current or potential musculoskeletal disorders is less well understood. Many clinicians assume the use of visual assessment is an effective way of diagnosing poor movement quality in this population and consequently prescribing an appropriate intervention. Therefore, the reliability and validity of visual observations in this group is essential and requires further investigation. Additionally studies using various clinical assessment tools often report reliability briefly as part of the main study without sufficient detail for appropriate interpretation. Kottner et al., (2011) suggested there is a need for reporting of reliability and agreement in separate publications with sufficient details as to study design and statistical analysis to make the data as useful as possible.

The visual assessment of lower extremity movement quality during functional tests is now widespread when assessing clients for risk of injury and during rehabilitation from current injuries. Visual assessment is considered a key component of the management of many lower extremity pain syndromes (Sahrmann, 2002). The double leg small knee bend (SKB) lower extremity functional test (also known as a partial or mini squat) and its variations (single leg SKB, lunge, hop lunge) are commonly used by physiotherapists and sports physicians to assess movement quality and dynamic lower extremity alignment. Poor dynamic alignment has been described as a combination of excessive pelvic drop, hip adduction and internal rotation, knee abduction, tibial internal or external rotation and foot hyperpronation (Earl, et al., 2005; Powers, 2003; Sahrmann, 2002). Assessment is focused on frontal and transverse plane control of the trunk, pelvis, hip and knee as this has frequently been linked to injury, particularly overuse problems such as patellofemoral dysfunction and iliotibial band syndrome (Powers, 2010; Reiman, Bolgla, & Lorenz, 2009).

In large scale prospective studies Boling et al., (2009) included increased hip internal rotation during a jump-landing task as a risk factor for developing patellofemoral dysfunction (PFD) and Willems et al., (2006) reported increased pronation as a risk factor for developing exercise related lower limb pain including shin splints, medial tibial stress, periostitis and stress reaction. In further prospective studies links have been shown between peak knee abduction and lateral trunk displacement and risk of non-contact ACL injury in females (Hewett et al., 2005; Zazulak, Hewett, Reeves, Goldberg, & Cholewicki,

2007). Retrospective studies using three-dimensional (3D) motion analysis have also reported a link between lower extremity alignment during functional tasks and PFD. During jumping tasks females with PFD had greater hip adduction excursion (Willson & Davis, 2009) and greater hip internal rotation (Souza & Powers, 2009). Hip adduction and internal rotation during a lateral stepdown have also been associated with PFD (Earl, Hertel, & Denegar, 2005). The association between hip control and PFD is further supported by two studies that reported deficits in hip strength and neuromuscular control in participants with PFD (Cowan et al., 2008; Ireland et al., 2003). Hip weakness may predispose female athletes to increased femoral adduction and internal rotation and thus greater risk of lower extremity injury (Leetun et al., 2004). While the link between hip strength and kinematics remains unclear, Reiman et al., (2009) concluded in a review that there was mounting evidence that hip weakness contributes to lower extremity injury across all ages. Retrospective studies using two-dimensional (2D) motion analysis have reported greater medial displacement of the knee (knee frontal plane projection angle) during a single leg squat in females with PFD compared to a control group (Levinger, Gilleard, & Coleman, 2007; Willson & Davis, 2008a). Overuse injuries other than PFD have also been linked to altered lower extremity alignment in female runners. Increased peak knee internal rotation and hip adduction angle has been reported in a group of female runners with a history of iliotibial band syndrome (Ferber et al., 2010; Noehren, Davis, & Hamill, 2007) and increased peak hip adduction and rearfoot eversion in female runners with previous tibial stress fracture (Milner et al., 2010).

Alongside assessment of movement quality clinicians often assume these functional tests are an effective method of diagnosing actual functional movement dysfunction (e.g. during running and landing), however the evidence for this appears mostly anecdotal. As well as being a common part of general clinical assessment the visual assessment of lower extremity movement quality has become a key component of screening athletes for risk of injury (Chiaia, et al., 2009; Reid, Stotter, Schneiders, Hing, & White, 2003). Recently assessment of lower extremity movement quality has been specifically recommended as a screening tool in young athletes (Ford, Myer, & Hewett, 2007; Örtqvist, et al., 2011). Hewett (2005) reported that prospective screening of dynamic alignment (three-dimensional kinematics) using a drop jump task could predict risk of knee injury in young female athletes from a range of sports. The increasing push for physical activity in youth as part of a healthy lifestyle makes

the risk of injuries an increasing concern and thus screening for risk factors has increased. Prevention of these injuries is crucial due to the cost and potential for injuries in youth to linger into adult years, resulting in early degenerative musculoskeletal disorders (Bahr & Holme, 2003). The development of valid functional tests to assess movement quality has the potential to aid injury prevention in this area.

The reliability and validity of visual assessment of lower extremity movement quality during the SKB has not been reported in the literature and the reliability and validity of other functional tests including the lunge, hop lunge and step-down needs further clarification. There has been limited attention given to the repeatability of an individual's performance in these functional tests, an important factor in establishing them as a valid clinical tool. The test protocol (e.g. variations in efforts to standardise range of knee flexion, velocity of the test) (Knudson, 2000; Zeller, McCrory, Kibler, & Uhl, 2003), the experience of the physiotherapist (Crossley, Zhang, Schache, Bryant, & Cowan, 2011) and the rating protocol (e.g. number of scoring categories, detail in rating criteria and amount of rater training) (Chmielewski, et al., 2007) may all affect the visual assessment outcome. Additionally the screening of lower extremity movement quality is specifically promoted in young athletes as an injury prevention strategy (Örtqvist et al., 2011). As young athletes' motor control and coordination is continuing to develop it can't be assumed young athletes will be able to repeat functional screening tests as well as adults or that the test results will have the same association to actual function. Furthermore due to the different anthropometric characteristics of the two groups it can't be assumed visual rating accuracy is generalizable and thus visual rating of children needs investigation as has been noted by others (Ekegren, Miller, Celebrini, Eng, & Macintyre, 2009).

The ability of visual assessments to identify groups with different kinematics as measured by two-dimensional (2D) and three-dimensional (3D) motion analysis has also received limited attention. The relationship between 2D and 3D kinematics during these functional tests also needs clarification as it may aid interpretation of visual assessment. Further development of these tests will improve the assessment and management of lower extremity musculoskeletal pain syndromes and aid the implementation of injury prevention programmes in the hope of minimizing lower extremity injuries in all ages.

Questions addressed in this PhD thesis

Given limitations in the literature, the overall question of this thesis was "Is physiotherapy visual assessment and kinematic assessment of dynamic alignment during lower extremity functional screening tests reliable and valid?" Specific questions were:

    I.   What is the typical variation in the kinematics of adults performing SKB tests, lunge, hop lunge and step-down and young athletes performing SKB and drop jump tests?

    II.   What is the association between kinematics during the functional tests and those during activities such as running and landing?

    III.   What is the reliability and validity of physiotherapist visual assessment of dynamic alignment during functional tests in adults and in young athletes?

    IV.   What is the influence of physiotherapist experience, assessment method and test velocity on visual assessment?

    V.   What is the association between 2D measures of alignment and 3D kinematics during functional tests in young athletes?

Structure of the PhD thesis

The thesis is structured into three thematic areas and nine chapters (see Figure 1.1). Theme 1: 'Visual and kinematic assessment of lower extremity functional screening tests' determined the context of the lower extremity functional screening test assessment issue, methods of assessment currently used by clinicians and researchers and knowledge of assessment reliability and validity. A review of literature and a technical note were completed to address the issues in Theme 1. The literature review was updated just prior to submission of the thesis in preparation for submission to *Sports Medicine* for potential publication. As a result the review includes results from this thesis and for this reason it has been included as the last chapter rather than earlier in the thesis which would be more common.

Theme 2: 'Kinematics' assessed the reliability of kinematics during lower extremity functional screening tests and the association with kinematics during actual function (running and landing). This consisted of two quantitative experimental cross-sectional studies using three-dimensional motion

analysis, one with healthy adults and another with healthy young athletes and resulted in three chapters.

Theme 3: 'Visual ratings' determined the reliability and validity of physiotherapy visual assessment of lower extremity functional screening tests. This consisted of two quantitative experimental cross-sectional studies, with a broad range of physiotherapists, assessing both adults and young athletes. Given the use of similar methods for various chapters, there is some overlap in chapters when reporting methodological details. Chapters two to nine within each of the three themes are outlined after Figure 1.1.

Physiotherapy visual assessment of dynamic alignment during lower extremity functional screening tests

Chapter 1: Preface/Introduction

Visual and kinematic assessment of lower extremity functional screening tests

Chapter 8: Visual and kinematic assessment of lower extremity dynamic alignment: A review of literature.
To be submitted to *Sports Medicine*

Chapter 2: Measures of agreement with visual assessment: What statistics should be presented?

Kinematics

Chapter 3: Kinematics during lower extremity functional screening tests – are they reliable and related to jogging?
***Physical Therapy in Sport*, 2011**

Chapter 5: Kinematics during lower extremity functional tests in young athletes – are they reliable and valid?
***Physical Therapy in Sport,* 2012, in review**

Chapter 7: The associations between 2D alignment and 3D kinematics during lower extremity functional tests in young athletes.
***Sports Biomechanics,* in review**

Visual rating

Chapter 4: Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests.
***Physical Therapy in Sport,* 2011**

Chapter 6: The reliability and validity of physiotherapist visual rating of dynamic pelvis and knee alignment in young athletes.
***Physical Therapy in Sport,* 2012, in review**

Chapter 9: Discussion/Conclusion

Appendices:
1. Whatman, C., Hing, W., & Hume, P. A. (2008, November). Physiotherapist agreement in rating movement quality during lower extremity functional tests *Sports Medicine New Zealand*. Symposium conducted at the meeting of the New Zealand Conference of Medicine and Science in Sport, Dunedin.
2. Whatman, C., Hume, P. A., & Hing, W. (2009, November). Lower extremity functional tests – are they reliable and do they predict actual function? *Sports Medicine New Zealand*. Symposium conducted at the meeting of the New Zealand Conference of Medicine and Science in Sport, Rotorua.

Figure 1.1: Overview of the structure of the thesis.

<u>Chapter 2</u> is a technical note providing a discussion of statistical methods used for reporting reliability/agreement when making visual assessments. We suggest overall percentage agreement (PA) and the First Order Agreement Coefficient (AC1) statistics should be presented for clinical populations where extremes of prevalence and/or rater bias are anticipated. These statistics for reporting reliability/agreement were later used in Chapters 5 and 7.

<u>Chapter 3</u> is a quantitative experimental cross-sectional study investigating the reliability of kinematics during lower extremity functional screening tests and the association between these kinematics and those occuring during running in adults. Peak three-dimensional trunk and lower extremity kinematics were quantified in 25 uninjured adults during five functional tests (SKB, single leg SKB, hop, lunge, hop lunge and step-down) and running. All functional tests were repeated by 10 adults one to two days later. For the majority of kinematic variables the within-day reliability was excellent (ICC ≥ 0.92) and the between-days reliability was excellent to good (ICC ≥ 0.80). The correlation between kinematics of the functional tests and running was generally large to very large (r = 0.53 to 0.93). Lower extremity functional screening tests should prove a useful clinical tool when assessing dynamic lower extremity alignment.

<u>Chapter 4</u> is a quantitative experimental cross-sectional study investigating physiotherapist agreement in visually rating movement quality/dynamic alignment during lower extremity functional screening tests performed by adults. Video recordings of six adults performing four lower extremity functional tests were visually rated (*dichotomous* and *ordinal* scale) using two rating methods (*overall* and *segment*) by 44 physiotherapists (33 physiotherapists repeated ratings 3-4 weeks later). Intra-rater agreement for overall and segment methods ranged from slight to almost perfect (PA: 29% to 96%, AC1: 0.01 to 0.96) and on average was moderate to good. AC1 agreement was better in the *experienced* group (84 to 99% likelihood) and for *dichotomous* rating (97 to 100% likelihood). Inter-rater agreement ranged from fair to good (PA: 45% to 79%; AC1: 0.22 to 0.71). AC1 agreement was not influenced by clinical experience but was again better using *dichotomous* rating. Rating of the knee position relative to the foot achieved the highest combination of intra-rater and inter-rater agreement of all the segments rated. Physiotherapists' visual rating of movement quality during lower extremity functional tests resulted in slight to almost perfect intra-rater agreement and fair to good

inter-rater agreement. Agreement improved with increased level of clinical experience and use of dichotomous rating.

Chapter 5 is a quantitative experimental cross-sectional study to determine if lower extremity functional tests are reliable and valid screening tests of lower extremity dynamic alignment in healthy young athletes. Peak three-dimensional pelvis and lower extremity kinematics were quantified in 23 uninjured young athletes during three lower extremity functional tests (SKB, Single Leg SKB and Drop Jump) and Running. All functional tests were repeated by 10 young athletes eight to ten weeks later. Within-day reliability was excellent (ICC ≥ 0.85) and between-days reliability was excellent to good (ICC range 0.60 to 0.92) for the majority of kinematic variables. Correlations for peak lower extremity kinematics between SKB and Drop Jump were moderate to very large (r = 0.39 to 0.87) as were correlations between Single Leg SKB and Running (r = 0.45 to 0.84). Small knee bend lower extremity functional tests are a useful clinical tool for assessing dynamic lower extremity alignment in healthy young athletes.

Chapter 6 is a quantitative experimental cross-sectional study investigating the ability of physiotherapists to visually rate dynamic pelvis and knee position in young athletes during lower extremity functional tests. Pelvis and knee alignment during lower extremity functional tests, in 23 young athletes, was visually rated by 66 physiotherapists (26 physiotherapists repeated ratings 3 to 4 weeks later). Physiotherapist ratings were compared to consensus visual ratings of an expert panel and the consensus ratings were also compared to peak 2D and 3D kinematics. Mean intra-rater agreement for all ratings was substantial (PA: 79% to 88%, AC1: 0.60 to 0.78). Inter-rater agreement ranged from fair to substantial (PA: 67% to 80%; AC1: 0.37 to 0.61). Sensitivity (≥80%) and specificity (≥50%) were acceptable for all tests except the Drop Jump. Experience (Diagnostic odds ratio 1.6 to 2.8 times better) and slower movement (4.9 times better) were possibly factors in better rating accuracy. Expert consensus ratings (aided by video slow motion) were able to differentiate young athletes with different 2D kinematics (very likely to almost certainly) and 3D hip kinematics (likely to very likely) but not 3D knee abduction. Visual rating by physiotherapists is a valid tool for identifying young athletes with poor frontal plane dynamic pelvis and knee alignment. Ratings are better with

slower movements and possibly with increased clinical experience. The finding that visual assessment is most closely related to 2D alignment measures provided the rationale for chapter 8.

Chapter 7 is a quantitative experimental cross-sectional study investigating the associations between 2D measures of alignment (knee frontal plane projection angle [FPPA] and lateral pelvic tilt) and 3D frontal and transverse plane pelvis, hip and knee kinematics in young athletes during lower extremity functional tests. Pelvis and lower extremity 3D kinematics at peak knee flexion were quantified in 23 uninjured young athletes during three lower extremity functional tests (Small Knee Bend [SKB], Single Leg SKB and Drop Jump). Lateral pelvic tilt and knee FPPA from 2D video were calculated using SiliconCoach7. There were moderate (r = 0.33 to 0.47) associations between knee FPPA and hip joint kinematics and small to moderate (r = 0.21 to 0.40) associations with knee joint kinematics. There were large to very large (r = 0.71 to 0.88) associations between knee FPPA and femoral and tibial segment frontal plane positions. There was a very large association between lateral pelvic tilt measured in 2D and 3D (r = 0.71, 90% CL, 0.56 to 0.81). Simple 2D measures of lower extremity alignment are useful for identifying kinematics that should be suspected of placing young athletes at risk of injury; however they should not be used to quantify 3D joint angles.

Chapter 8 is a review of the literature to investigate the evidence for visual and kinematic assessments of movement quality/dynamic alignment during lower extremity functional screening tests. This review was originally drafted in 2007 (in the earlier stages of the PhD) and there was little evidence for the reliability and validity of visual assessment of several lower extremity functional screening tests, in particular the small knee bend (SKB), lunge and hop lunge. At that time the information in the review formed the premise for Chapters 4, 5, 6 and 7. The review was subsequently updated to include results from this thesis in preparation for submission to *Sports Medicine* for potential publication.

Chapter 9 is an overall discussion of the key findings, implications and limitations of the preceding chapters and areas for further research.

The appendices contain material from chapters 4 and 5 that were presented as conference presentations (see Appendices 1 and 2). Subject information sheets and consent forms are provided in Appendix 7 and 8. Appendix 9 contains notifications from the Auckland University of Technology Ethics Committee (AUTEC) regarding ethical approval for the studies.

CHAPTER 2: MEASURES OF AGREEMENT WITH VISUAL ASSESSMENT: WHAT STATISTICS SHOULD BE PRESENTED?

Overview

Several statistical methods are discussed in the literature for reporting agreement between two or more clinicians who have made independent visual assessments using categorical ratings. The focus of this technical note are situations where trait prevalence is likely to be high or low, considered common in many conditions presenting to physiotherapists and likely to occur in screening situations with healthy individuals. The simplest (best understood) measure of agreement is the overall percent agreement however it is criticised for failing to account for agreement by chance. As a result several agreement coefficients have been proposed that are said to correct for the influence of chance agreement, the most popular being the Kappa statistic. However due to the limitations of Kappa several other agreement coefficients have been proposed including the first-order agreement coefficient (AC1). The AC1 also adjusts overall agreement for chance agreement but is said to be more resilient to the problems of trait prevalence and rater bias that affect Kappa. The AC1 is suggested to be more consistent with percent agreement in all situations. We suggest overall percentage agreement and the AC1 statistics should be presented for clinical populations where extremes of prevalence and/or rater bias are anticipated.

Methods

Following consultation with a statistician a directed search was made of articles and websites discussing statistical methods for reporting agreement among raters using dichotomous and ordered categorical ratings. All relevant sources were reviewed for recommendations on how to most appropriately measure agreement with a focus on situations of extremes of trait prevalence and/or rater bias.

Findings

Several methods are described in the literature for studying agreement among ratings made by two or more raters. Hayes and Krippendorff (2007) reported up to 43 measures for estimating the reliability of nominal data and thus a full discussion is beyond the scope of this chapter. Rather a discussion of key points to consider when estimating agreement of categorical ratings, including dichotomous and ordered categorical ratings is presented. Of most interest are situations where trait prevalence is likely to be high or low, considered common in many conditions presenting to physiotherapists (Fritz & Wainner, 2001) and likely to occur in screening situations with healthy individuals.

Given the importance of rater agreement to clinical decision making there is surprisingly little agreement in the literature as to the most appropriate statistics to present. The choice seems to depend on the nature of the underlying data and the opinions of authors on the assumptions and use of various probability models employed to adjust observed levels of agreement. Uebersax (n.d.) suggested less complicated statistical methods were preferable as they reveal more "common sense" information about agreement in the data and have more inherent clinical meaning. He added that in the majority of cases, advanced methods based on models of observed and chance agreement, should be used as complements, not substitutes for simpler methods.

The simplest measure of agreement appears to be the overall percent agreement. This is calculated by expressing the number of observed agreements as a proportion of maximum possible agreements. The level of percent agreement appropriate for clinical use does not appear to have been given much consideration in the literature. This is likely because, although it has the maximum common sense value and seems the most easily interpreted by clinicians, taken by itself it has proposed limitations. Overall percent agreement doesn't distinguish between agreement on positive ratings versus agreement on negative ratings and it is also often criticised for failing to account for agreement by chance. In the opinion of many authors this raises the concern that the proportion of overall agreement is prone to chance related inflation or bias as it includes agreements that do not reflect true intentional agreement between the raters (Sim & Wright, 2005).

Due to the concerns around overall percent agreement several agreement coefficients have been proposed that are said to correct for the influence of chance agreement. The most popular are the kappa and weighted kappa statistics which are said to give overall agreement beyond that expected by chance - considered the true agreement (Sim & Wright, 2005). Weighted kappa is said to be more appropriate for ordered categorical data. It reflects the degree of disagreement by placing greater emphasis on large differences between ratings than on small differences. Although appealing, the weightings applied seem quite arbitrary and different weighting schemes produce different values of weighted kappa for the same data (Sim & Wright, 2005). Although kappa is common researchers have been criticised for using it solely for this reason without consideration of issues such as extremes of trait prevalence and rater bias (Uebersax, n.d.).

The paradox of kappa, where high agreement is coupled with a low kappa value (Feinstein & Cicchetti, 1990) has received attention in the literature. Several authors have highlighted this limitation among others associated with kappa (Blood & Spratt, 2007; Di Eugenio & Glass, 2004; Gwet, 2008a; Sim & Wright, 2005). The paradox appears to exist when trait prevalence (the number of cases in the population) is very high or low. This may occur in physiotherapy screening situations so the kappa statistic in these instances may be artificially lowered and difficult to interpret. Overall agreement may be high, but kappa will be low and will thus not reflect true agreement between raters. Gwet (2008a) suggested that kappa only gives a reasonable measure of agreement when trait prevalence is close to 50% and its ability to reflect agreement diminishes considerably as prevalence gets closer to 0 or 100%. He argued that the model used to estimate chance agreement in the kappa calculation is inappropriate, under high or low prevalence conditions, as it overstates the correction due to chance agreement. Hoehler (2000) suggested the use of kappa should be restricted to studies with a prevalence of approximately 50%. Vach (2005) disagreed stating that we should worry less about the influence of prevalence on kappa and be more concerned about other attributes of the sample composition. He also suggested that to aid interpretation and comparison of kappa between studies, authors need to describe the fraction of subjects who were difficult to decide on. It is however not clear from the article how this fraction would be achieved in a variety of situations.

Kappa is also highly sensitive to rater bias as indicated by differences in the rater marginal probabilities (Gwet, 2002; Sim & Wright, 2005). Bias indicates the extent to which the raters agree or disagree regarding the proportion of cases in each category. Gwet (2008a) suggested that differences in the rater marginal probabilities contributes to the unstable nature of the kappa statistic and adds to the difficulty of its interpretation. Rater bias has been termed the second paradox of kappa (Sim & Wright, 2005) and results in higher kappa values when bias is large. It has been suggested that in situations of high trait prevalence and rater bias, kappa fails the common sense test and is an inappropriate measure of agreement (Haley, 2007). Gwet (2008b) suggested that although the seriousness of the paradoxes of kappa have been minimised by some authors they remain a major problem for practitioners. We acknowledge there are adjustments that can be made to kappa to account for the influence of prevalence and bias giving the Prevalence-Adjusted Bias-Adjusted Kappa (PABAK) (Sim & Wright, 2005). This seems to possess even less common sense value (clinical meaning) with further corrections based on additional hypothetical situations and there is further disagreement in the literature on the use of this PABAK statistic.

Additionally not all authors agree with the logic of a chance corrected agreement coefficient in the first place. Uebersax (n.d.) suggested the logic of comparing an actual observed level of agreement with a hypothetical chance level, which may occur under an unrealistic model, is unclear. Furthermore he suggested that it is not clear how chance affects the decision of actual raters and how one might correct for it. He argued the need for chance correction has not been convincingly demonstrated and the assumptions underlying the correction (as calculated in the kappa statistic) are flawed. Uebersax (n.d.) pointed out that one assumption underlying the estimate of chance agreement in the kappa statistic is that raters guess on every rating which we agree is completely untenable. From a practical view point clinicians may struggle to understand why chance agreements (that may or may not occur and that you cannot remove from clinical practise) would be removed from reported agreement in studies.

Uebersax (n.d.) further argued against correcting for chance when measuring agreement given we don't do so when we measure the accuracy of a diagnostic test by reporting sensitivity. He pointed out sensitivity is not corrected for chance probably because specificity is generally also reported. In a

situation of high prevalence and lots of positive test results, sensitivity will be high even if the test and the true diagnosis are independent. Consideration of low specificity (alongside the high sensitivity) however means the test is viewed sceptically and avoids the need for chance correction. Rather than adjusting for chance agreement (which must be estimated based on questionable assumptions), Uebersax (n.d.) suggested a joint consideration of the specific proportion of positive and negative agreement. He argued that if both are sufficiently large (analogous to considering sensitivity and specificity) there is no need for adjusting for chance as overall agreement is less liable to chance related inflation or bias. Although we tend to agree with the view expressed by Uebersax (n.d.) we could not find this approach used in relevant literature. This may be because although the calculations are simple they are not readily available for large data sets in commonly used statistical software.

Due to the limitations of Kappa two further agreement coefficients have recently been proposed Gwet (2008a). The first- and second-order agreement coefficients AC1 and AC2 also adjust overall agreement for chance agreement but are said to be more resilient to the problems of trait prevalence and rater bias. Gwet (2008a) stated the AC1 statistic uses a chance-agreement probability that overcomes the two paradoxes of kappa discussed earlier. Haley (2007) presented an example that supports the claim that the AC1 is a more robust statistic that yields reliable results. AC1 estimates chance agreement based on the portion of ratings that may lead to an agreement by chance, rather than assuming all observed ratings may yield an agreement by chance, as is calculated for the kappa statistic. In addition to adjusting more appropriately for chance agreement, as for AC1, the AC2 statistic additionally adjusts for misclassification errors (Blood & Spratt, 2007).

Gwet (2002) proposed there are two components to agreement, the assessment of the presence of the trait and then the classification into a particular category on some rating scale. Uebersax (n.d.) also commented that both of these components of agreement should be considered where possible. Gwet (2002) devised the AC2 statistic to adjust for misclassification, a situation he suggested may exist where two raters agree on the assessment of a trait but classify the participant into different categories. If this adjustment for misclassification is not made Gwet (2002) suggested that inter-rater agreement will be inappropriately lowered. Although appealing, this adjustment for misclassification in the AC2 relies on the estimation of conditional misclassification probabilities which Gwet (2002) stated

are best estimated based on prior reliability studies using the same scale. It seems obvious that this would often not be the case. It is also not immediately clear that breaking agreement into two components makes sense in all situations. Given these concerns with the AC2 it may be more appropriate at this stage to use the AC1 for ordered categorical scales although it may give a conservative estimate of true agreement in this type of data. One of the likely reasons the AC1 and AC2 statistics have not been commonly used is because their calculation is not readily available in common statistical software packages. However, several authors in a variety of disciplines have begun to use the AC1 statistic (Cheng, Laron, Schiffman, Tang, & Frishman, 2007; Kernaghan & Penney, 2006) and macros for its calculation have been made available for both the statistical packages SAS and SPSS. Two researchers at the Dartmouth Medical School have recommended the use of the AC1 and made a macro for the statistical package SAS freely available (Blood & Spratt, 2007). They do however caution that it is a new statistic, needing further examination, prior to it being adopted as the standard. Gwet (2001) suggested the AC1 can be interpreted in a similar manner to the kappa coefficient, based on a scale proposed by Landis and Koch (1977): 0.01-0.20=slight; 0.21-0.40=fair; 0.41-0.60=moderate; 0.61-0.80=good; 0.81-1.0=almost perfect.

For its intuitive, common sense value to therapists it appears that the proportion of overall agreement is an important agreement statistic and it could be reported alongside the proportions of positive and negative agreements. However presenting both positive and negative agreement values makes it difficult to compare the effect on agreement of factors such as rater experience as raters may perform better on one but not both. Furthermore the current view still suggests that chance agreement must also be considered to indicate true agreement in keeping with the original proposal of Cohen (1960). Increasingly the AC1 statistic is gaining support as the most robust statistic for this purpose across a range of all values of trait prevalence. Gwet (2008a) has shown the AC1 behaves in a similar manner to kappa when trait prevalence is in the vicinity of 50%, but outperforms kappa when it moves towards the extremes. Others have acknowledged the AC1 is more consistent with percent agreement in all situations and have recommended its use (Chan, 2003). We suggest that the percentage agreement and AC1 statistics should be presented for clinical populations where extremes of prevalence and/or rater bias are anticipated.

CHAPTER 3: KINEMATICS DURING LOWER EXTREMITY FUNCTIONAL SCREENING TESTS – ARE THEY RELIABLE AND RELATED TO JOGGING?

Overview

Purpose: To investigate the within-day and between-day reliability of 3D lower extremity kinematics during five lower extremity functional screening tests and to assess the association between these kinematics and those recorded during jogging.

Methods: Peak three-dimensional lower extremity kinematics were quantified in 25 uninjured participants during five lower extremity functional tests and jogging. A nine camera motion analysis system (Qualysis Medical AB, Sweden) was used to capture three trials of all tests. All functional tests were repeated by 10 participants one to two days later. Visual 3D (C-Motion Inc, USA) and Labview were used to process all data. Intraclass correlation coefficients (ICC) and typical errors (TE) were used to assess within- and between-day reliability of all variables. Pearson correlation coefficients were used to evaluate the association between peak joint kinematics during the functional tests and jogging.

Results: For the majority of kinematic variables the within-day reliability was excellent (ICC ≥ 0.92) and the between-day reliability was excellent to good (ICC ≥ 0.80). The correlation between kinematics of the functional tests and jogging was generally large to very large (r = 0.53 to 0.93).

Conclusions: These results suggest these lower extremity functional screening tests should prove a useful clinical tool when assessing dynamic lower extremity alignment.

Background

Screening of individuals for risk of future injury and as a means to optimising performance has become common, particularly in professional sport but also at other competitive and recreational levels (Mottram & Comerford, 2008). When screening the lower extremity the use of functional tests to evaluate movement quality (neuromuscular control) is now highly recommended. Functional tests are frequently used to identify altered lower extremity kinematics during weight bearing activities in the belief this is linked to injury risk and peak performance. However it must be acknowledged the validity

of these tests for predicting injury and/or performance remains unclear. Despite this traditional assessment of isolated joints and muscles alone is no longer considered adequate and the additional use of functional tests to assess multiple muscles, across multiple joints, in functionally orientated tasks is common (Mottram & Comerford, 2008). Physiotherapists utilise this information (primarily gained from visual observation) in their clinical decision making process when considering their prescription of exercises and also evaluating progress during rehabilitation.

The Small Knee Bend (SKB) (Mottram & Comerford, 2008; Reid, et al., 2003; Sahrmann, 2002) and its variations (double leg, single leg, lunge, hop lunge) is a common lower extremity functional test used by physiotherapists and sports physicians to assess dynamic trunk and lower extremity alignment. This test (also known as a partial squat) is described by Sahrmann (2002) as part of a lower quarter examination. A single leg SKB has also been described as a specific screening test by Mottram and Comerford (2008). Additionally it is common clinical practise, when indicated, to further evaluate dynamic trunk and lower extremity alignment by using a lunge (Crossley, et al., 2006) and hop lunge (Cook, 2006). Similar tests reported in the literature include single leg squats (SLS) (Zeller, et al., 2003) and single leg step downs (Earl, et al., 2007).

When using lower extremity functional tests, such as those described above, physiotherapists evaluate dynamic trunk and lower limb alignment. Poor dynamic alignment has been described as a combination of excessive trunk lateral flexion, pelvic drop, hip adduction and internal rotation, knee abduction, tibial internal or external rotation and foot hyperpronation (Earl, et al., 2005; Powers, 2003; Sahrmann, 2002; Willson & Davis, 2009). The resultant excessive medial displacement of the knee in the frontal plane has also been termed dynamic knee valgus (Bell, Padua, & Clark, 2008). The movements of concern to clinicians mostly occur in the frontal or transverse planes. Clinical observation of the position of the patella relative to a line extending vertically from the 1st or 2nd toes is also a common measure of frontal plane control (Bell, et al., 2008; Hirth & Padua, 2007; Mottram & Comerford, 2008). Abnormal motion of the trunk, pelvis, hip and knee in these planes, observed during activities such as running, squatting and landing, is considered a key risk factor for the development of common injuries such as patellofemoral dysfunction (Powers, 2003).

Few studies have investigated the reliability of trunk, pelvic, hip, knee and ankle 3D joint kinematics during lower extremity functional tests. We are not aware of any studies reporting kinematic reliability during a SKB, lunge or hop lunge or any studies reporting the reliability of trunk and pelvic kinematics during any tests including the common SLS. Two studies have used 2D techniques to assess the reliability of frontal plane kinematics during a SLS (Levinger, et al., 2007; Willson, et al., 2006) reported the measure was reliable within-day (ICC=0.88), while (Levinger, et al., 2007) reported acceptable reliability within and between-days (ICC 0.88 and 0.74). Neither study reported an absolute measure of reliability. In addition to the lack of studies investigating the reliability of SLS, the generalisability of these to SKB is questionable. Although similar, and looking to assess the same alignment faults, there are subtle differences in the performance of SKB and SLS. The focus of a SLS is most often the flexion of the hip with associated trunk flexion. The SKB emphasises the flexion of the knee and ankle (while maintaining a relatively upright trunk position) in a pattern that possibly more closely simulates the stance phase of walking and running. Furthermore to standardise the performance of SLS, previous studies have relied on monitoring the amount of knee flexion (Claiborne, et al., 2006; Levinger, et al., 2007; Willson, et al., 2006). This is not common clinically due to the extra time and equipment required. This combined with possible differences in the movement pattern make any reliability reported less generalisable to clinical tests such as SKB. The reliability of hip and knee joint 3D kinematics (but not trunk and pelvis) during a stepdown and lateral stepdown has also been reported to be acceptable (ICC= 0.70 to 0.88) (Bolgla, et al., 2008; Earl, et al., 2005). Within-subject kinematic variation during functional tests is one important type of reliability for physiotherapists to consider if they want to visually assess and monitor performance of their clients (Hopkins, 2000). It is crucial to know if the kinematics are consistent enough from day to day for making clinical decisions. For example, a physiotherapist needs to know if following a rehabilitation programme the change in hip adduction noted visually during a lower extremity functional test is real or whether the change is due to expected kinematic variability with repeat testing. The reporting of absolute variability (such as typical error in degrees) also provides greater clinical meaning than the more commonly reported relative variability (such as ICC). Furthermore the reliability of these tests also needs to be established if they are to be used in longitudinal studies evaluating injury risk or the effect of rehabilitation interventions.

An advantage of functional testing is that it is thought to replicate the kinematics encountered during a task specific activity (Clark, 2001). The SKB test and its various extensions are considered useful in the clinic to gain an insight into the kinematics a client may exhibit during functional tasks such as walking, running, stair climbing and lunging. A recent study by Willson and Davis (2008a) suggested a link between kinematics (2D frontal plane projection angle) during a single leg squat and more dynamic activities (running and jumping). These authors found that a group with patellofemoral pain had less internal hip rotation when squatting, jumping and running than a control group. Clinicians often assume these functional tests are an effective method of diagnosing actual functional movement dysfunction and consequently prescribing prevention or rehabilitation programmes. However the evidence for this is mostly anecdotal. Therefore the association between the kinematics recorded during SKB tests and those occurring during actual function (e.g. jogging) need investigation. This relates to the validity of the tests and when clinicians select a test they need to acknowledge issues relating to both reliability and validity (Clark, 2001).

To date, no research has reported the typical variation in trunk, pelvis, hip, knee and ankle 3D joint kinematics a participant would exhibit with various SKB tests within or between days. The use of SKB type tests as a clinical screening tool for walking and running gait also needs investigation.

Purpose

To investigate the within- and between-day reliability of peak 3D trunk and lower extremity kinematics during five lower extremity functional screening tests and to assess the association between these kinematics and those recorded during jogging. Peak kinematics in the transverse and frontal plane during loading were chosen as they are frequently screened in the clinic and linked to risk of injury.

Methods

Participants

Twenty five participants (mean age = 22 ±4 yr, mean height = 171 ±10 cm, mean weight = 66 ±12 kg) with no musculoskeletal problems volunteered for this study. The study was approved by the

Auckland University of Technology Ethics Committee. All participants received verbal and written information about the study and gave written informed consent prior to testing.

Instrumentation

A University Motion Analysis Laboratory was used for all testing. This laboratory contains a nine camera motion analysis system (Qualysis Medical AB, Sweden) suitable for recording whole body 3D kinematics. Cameras, sampling at a rate of 240 Hz, were positioned to provide the optimum field-of-view of the area of the laboratory used for all testing. Prior to data collection the system was calibrated as per the manufacturer's protocol using a static calibration frame (to orientate the cameras with respect to the laboratory coordinate system) and a dynamic wand. During calibration with the dynamic wand average movement residue (RES) for the retro-reflective markers was less than 2 mm.

Prior to testing the trunk, pelvis and dominant leg of all participants were instrumented with 15 retro-reflective markers (19 mm diameter) secured to specific anatomical locations (bilateral acromion processes, bilateral ASIS's, sacrum, bilateral iliac crests, medial and lateral femoral epicondyles, mid-patella, medial and lateral malleoli, head of 5th metatarsal, head of 2nd metatarsal, posterior calcaneus). All markers were positioned by an experienced musculoskeletal physiotherapist based on palpation of appropriate anatomical landmarks (Figure 3.1).

Figure 3.1:   Participant instrumented with markers (note medial knee and ankle anatomical markers used for skeletal model construction are not shown as they were removed during functional tests).

These anatomical markers provided a reference marker set for construction of a skeletal model using a commercial biomechanical analysis software programme (Visual 3D, C-Motion Inc, USA) To ensure optimal reproduction of marker placement during repeat testing, a marker pen was used to identify marker position and participants were asked to retain this until the repeat testing session conducted within two days of the initial test. The dominant leg was identified by asking each participant the leg they would use to kick a ball.

Two cluster marker sets (a group of four retro-reflective markers attached to a light weight rigid plastic shell) were also attached to the thigh and shank of the dominant leg (Figure 3.1). These marker sets were designed to track motion of the thigh and shank segments. It has been suggested that clusters are more accurate and practical for tracking motion than individual skin markers (Angeloni, Cappozzo, Catani, & Leardini, 1993) and four markers attached to a rigid shell is thought to be optimal (Cappozzo, Cappello, Della Croce, & Pensalfini, 1997; Manal, McClay, Stanhope, Richards, & Galinat, 2000). All anatomical markers and the cluster marker sets were attached directly to the skin

with double sided adhesive tape. The cluster marker sets were additionally secured with elasticized Velcro straps of various lengths designed specifically for this purpose.

Testing protocol

All twenty five participants attended the Motion Analysis Laboratory on one occasion. Ten participants (40%) returned on a second occasion for repeat testing one to two days later. Following instrumentation of the retro-reflective markers a static trial was first collected in which participants were asked to stand still with foot placement standardized to the laboratory coordinate system.

Performance of the five functional tests was then randomized among participants. All participants were given standardized verbal instructions prior to each test (Table 3.1) and the researcher demonstrated each test. Participants were required to keep their heels on the ground throughout each test (except the step-down) in order to try and standardize the range of hip and knee flexion without the need for additional monitoring and equipment. These simple instructions increase the clinical utility of the tests. Practice for all tests was allowed until the researcher was confident the test was performed consistently (this usually required 3-5 practice attempts). All tests were performed within the same area of the laboratory and the velocity of each test was set by a three second count made by the researcher (verbally counted as "one and two and three") during the dorsiflexion/knee flexion phase of each test. Prior to the SKB and single leg SKB tests subjects walked three steps to move into the test area of the laboratory and take up their natural stance position prior to commencing each test. During all tests participants were instructed to maintain visual focus on a cross positioned on a wall directly in front of them at eye level. Pilot testing showed this to be a useful method for maintaining an upright trunk position, improving consistency and it was also thought to most appropriately simulate a functional head and trunk position. Participants performed three repetitions of each test and were also recorded jogging the length (~10 m) of the laboratory. Jogging trials were repeated until the researcher was convinced that three consistent trials had been collected (this usually required approximately five attempts). Mean jogging velocity, estimated from the anterior velocity of the sacral marker in the laboratory coordinate system, was 2.9 ±0.4 m/s.

Table 3.1:   Description of the five functional tests used in the study.

Functional test	Test description
Small knee bend (SKB)	Starting from a standing position participants performed a partial squat (hip and knee flexion) with the trunk maintained in an upright position. Participants were instructed to continue the SKB until they reached maximum dorsiflexion without lifting their heels and then return to upright standing.
Single leg small knee bend (dominant leg)	Standing on the dominant leg only, with the contralateral hip in neutral and knee flexed to approximately 80°, participants performed a SKB as described above.
Lunge (dominant leg)	From a standing position participants were instructed to lunge forward (leading with their dominant leg) a distance of approximately one and a half times the length of their normal gait stride. As they moved into single leg stance (on the dominant leg, with the contralateral leg off the ground) they flexed the hip and knee while maintaining an upright trunk. Participants were instructed to continue the lunge until reaching maximum dorsiflexion of the stance leg without lifting their heel.
Hop lunge (dominant leg)	From a standing position participants were instructed to jump forward a distance of approximately 1.0 m and on landing on the dominant leg to flex the hip and knee. Participants were instructed to continue the lunge until reaching maximum dorsiflexion of the dominant leg without lifting their heel.
Step-down	Leading with the non-dominant leg, participants stepped slowly down from a step 20 cm high.

Data processing

All static and motion (functional test and jogging) trails were tracked using the Qualysis motion capture software and exported to Visual 3D (C-Motion Inc, USA). In Visual 3D the static trial, combined with the height and weight data of each participant was used to create geometric objects of appropriate shape and mass to represent each body segment (trunk, pelvis, thigh, shank and foot). All lower limb segments were modeled as frusta of right cones and the pelvis and trunk segments were

modeled as right elliptical cylinders. Together these segments formed a 6-degree-of-freedom, rigid link biomechanical model. Visual 3D defines joints in the model as places where the distal end of one segment meets the proximal end of another segment and analysis of joint motion is based solely on the relative motion between the segments.

In Visual 3D the rigid link model created from the static file was then assigned to all the imported motion files to allow calculation of relevant kinematic data. The data from the motion files was filtered with a second-order Butterworth bidirectional low-pass filter with a cut-off frequency of 6 Hz. The Cardan sequence y-x-z was used for the calculation of joint angles which was equivalent to flexion/extension, abduction/adduction, axial rotation and in this case equivalent to the Joint Coordinate System described by Grood and Suntay (1983). The only exception to this was for calculation of pelvic angle with respect to the laboratory where we used the sequence z-x-y (axial rotation, obliquity, tilt) which is recommended in the Visual 3D documentation based on the suggestions of Baker (2001). Joint angles were not normalised to the static standing trial.

All kinematic data for the trunk, pelvis and lower limb were exported as 'text' files for importing into Labview for further analysis. The Labview VI processed the trunk, pelvis and lower limb, kinematic data and output the peak joint angles during the knee flexion phase of each functional test and the stance phase (start to maximum knee flexion only) of jogging. The mean of each of these variables, across the three repetitions of each functional test and jogging was then used in the between-day statistical analyses. We also output the maximum medial position of the patella marker in the frontal plane (y coordinate in the laboratory coordinate system) We have termed this medial knee displacement (MKD) and analysed it in the same manner as for the joint angles described above.

Statistical analyses

Proc mixed in Statistical Analysis Systems (SAS), "Version 9.1, SAS Institute, Cary, NC" was used to estimate within subject trial to trial variability within a day and between days. The trial to trial variability within a day was the residual error in the model. The model allowed estimation of a different residual for day 1 and day 2. The model included the random effect of Subject by Day. The model also included fixed effects (interaction of Day and Trial) to allow for a change in the mean between all trials

on both days. The mixed model allowed calculation of within- and between-day reliability expressed as a typical error (TE) in degrees and Intra-class correlation coefficient (ICC) with 90% confidence limits. The ICC we calculated in SAS is equivalent to an ICC (2, 1). Typical error is interpreted as the expected variation in peak joint angle when one individual is tested on repeat occasions. Additionally processing of errors in SAS was completed to estimate the ICC and typical error for different numbers of trials. The ICC classifications of Fleiss (1999) were used to describe the magnitude of ICC values (less than 0.4 was poor, 0.4 to 0.75 was fair to good, and greater than 0.75 was excellent).

Using the data from initial testing only (all twenty five participants), Pearson correlation coefficients were calculated to assess the magnitude of the association between peak joint angles during the functional tests and those during jogging. The magnitudes of these correlations were described as trivial (0.0-0.1), small (0.1-0.3), moderate (0.3-0.5), large (0.5-0.7), very large (0.7-0.9), or extremely large (0.9-1.0) (Hopkins, Marshall, Batterham, & Hanin, 2009). Using this same data mean peak joint angles and standard deviations have been calculated for each of the functional tests.

Results

The mean joint angles recorded during the knee flexion phase of each functional test are described in Table 3.2. Angles appear similar across all tests with a few notable exceptions. The step-down test recorded a much higher knee flexion angle (91°) than the other tests and SKB recorded a relatively low peak hip adduction angle (2°). On visual inspection of the data most peak angles in the frontal and transverse planes generally occurred between mid-range and maximum knee flexion.

Table 3.2:   Peak angle (°) during the knee flexion phase of each functional test for three trials (all 25 participants, mean ±SD).

		Hop Lunge	Single Leg SKB	SKB	Step-down	Lunge
Ankle	Dorsiflexion	17 ±5	19 ±7	17 ±7	17 ±7	19 ±7
	Eversion	16 ±5	10 ±4	13 ±3	16 ±5	13 ±5
Knee	Flexion	68 ±9	65 ±13	68 ±15	91 ±7	68 ±12
	Abduction to adduction*	3 ±4 to 9 ±5	0 ±3 to 9 ±5	0 ±3 to 7 ±5	0 ±3 to 11 ±5	1 ±4 to 8 ±5
Hip	Flexion	46 ±12	34 ±18	40 ±18	31 ±10	41 ±14
	Abduction to adduction*	-3 ±3 to 9 ±5	4 ±3 to 11 ±5	-2 ±4 to 2 ±3	1 ±4 to 10 ±4	3 ±4 to 13 ±6
	External to internal rotation[†]	-3 ±6 to 4 ±5	0 ±6 to 5 ±7	0 ±6 to 4 ±6	-1 ±5 to 5 ±6	0 ±6 to 6 ±7
Pelvis	Lateral tilt (L to R)	-4 ±4 to 1 ±3	-2 ±4 to 3 ±3	-2 ±2 to 0 ±2	-6 ±3 to 3 ±3	-4 ±4 to 1 ±3
	Transverse plane rotation (R to L)	0 ±4 to 9 ±4	-1 ±4 to 3 ±3	-1 ±2 to 2 ±3	-5 ± 4 to 1 ±4	1 ±4 to 7 ±5
Trunk	Lateral tilt (L to R)	-1 ±3 to 4 ±3	1 ±2 to 4 ±2	-1 ±1 to 0 ±1	-1 ± 2 to 3 ± 2	-1 ±2 to 2 ±2
	Transverse plane rotation (R to L)	-2 ±4 to 3 ±4	-1 ±4 to 2 ±4	-1 ±3 to 1 ±3	-4 ± 4 to 2 ±4	0 ±4 to 3 ±3

*range is minimum to maximum where +ve = adduction.
[†]range is minimum to maximum where +ve = internal rotation.

Based on the use of three trials, the within-day ICC values for all measured variables (ICC = 0.79 to 1.0; Table 3.3) are generally higher than the between-day ICC values (ICC = 0.46 to 0.99; Table 3.4). Within-day ICC's were all greater than 0.90, except for trunk lateral flexion during Hop Lunge and Lunge (ICC = 0.79). Within-day typical errors for all variables, representing the typical variation for an individual participant on repeat tests, ranged from 0.5 to 1.3°. The within-day ICC and typical error range across all tests was small indicating consistent reliability irrespective of the test.

Table 3.3: Within-day reliability of the peak angle (º) and medial knee displacement (cm) for all five functional tests

		Mean Typical Error (range)		ICC range	
Number of trials		3	10	3	10
Trunk	Lateral flexion	0.9 (0.3-1.6)	0.5 (0.2-0.9)	0.79-0.93	0.92-0.98
	Rotation	0.8 (0.5-1.1)	0.4 (0.3-0.6)	0.92-0.97	0.98-0.99
Pelvis	Lateral tilt	0.8 (0.6-1.0)	0.4 (0.3-0.5)	0.94-0.98	0.98-0.99
	Rotation	0.8 (0.6-1.0)	0.4 (0.3-0.5)	1.0-1.0	1.0-1.0
Hip	Flexion	1.3 (0.97-1.8)	0.7 (0.5-1.0)	0.99-1.0	1.0-1.0
	Adduction	0.8 (0.4-1.1)	0.5 (0.2-0.6)	0.97-0.99	0.99-1.0
	Internal rotation	0.7 (0.4-0.8)	0.4 (0.2-0.5)	0.99-1.0	1.0-1.0
Knee	Flexion	1.3 (1.1-1.7)	0.7 (0.6-0.9)	0.94-0.99	0.98-1.0
	Abduction	0.5 (0.4-0.9)	0.3 (0.2-0.5)	0.93-0.99	0.98-1.0
Ankle	Dorsiflexion	0.7 (0.5-1.0)	0.4 (0.3-0.5)	0.97-1.0	0.99-1.0
	Eversion	1.0 (0.7-1.4)	0.5 (0.4-0.8)	0.94-0.99	0.98-1.0
Knee	MKD	0.5 (0.2-0.7)	0.3 (0.0-0.4)	0.94-0.98	0.98-1.0

Typical Error 90% CL ~×/÷1.20, lowest ICC 90% CL ~ ±0.14

The majority of the between-day ICC's were greater than 0.8 and the typical errors ranged from 1.2 to 3.9°. The poorest between-day reliability was for trunk lateral flexion during the Single Leg SKB (ICC = 0.46), however the typical error was still only 1.2°. The trunk generally showed the worst between-day reliability. The hip showed the most consistent between-day reliability in all planes (ICC's = 0.87 to 0.98).

The additional processing of the errors did not suggest substantial improvements in within or between-day reliability if the number of trials was increased to ten (Table 3.3 and 3.4).

Table 3.4:   Between-day reliability of the peak joint angle (°) and medial knee displacement (cm) for all five functional tests

		Mean Typical Error (range)		ICC range	
Number of trials		3	10	3	10
Trunk	Lateral flexion	1.2 (0.4-1.7)	0.9 (0.4-1.6)	0.46-0.84	0.51-0.92
	Rotation	2.0 (1.8-2.6)	1.5 (1.1-2.2)	0.61-0.86	0.62-0.89
Pelvis	Lateral tilt	1.4 (0.9-2.1)	1.4 (0.8-2.0)	0.70-0.89	0.72-0.92
	Rotation	1.6 (1.1-1.9)	1.4 (1.0-1.8)	0.90-0.99	0.90-0.99
Hip	Flexion	3.9 (2.8-5.4)	3.7 (2.7-5.2)	0.90-0.98	0.91-0.98
	Adduction	1.9 (1.0-2.4)	1.8 (1.0-2.3)	0.87-0.97	0.88-0.98
	Internal rotation	3.1 (2.8-3.3)	3.0 (2.8-3.3)	0.91-0.97	0.91-0.97
Knee	Flexion	2.8 (2.4-3.7)	2.7 (2.2-3.6)	0.86-0.97	0.89-0.97
	Abduction	1.9 (1.7-2.3)	1.8 (1.7-2.3)	0.67-0.76	0.68-0.77
Ankle	Dorsiflexion	2.3 (2.0-2.4)	2.2 (1.9-2.4)	0.87-0.90	0.88-0.91
	Eversion	2.4 (1.4-2.7)	2.2 (1.2-2.6)	0.62-0.94	0.64-0.96
Knee	Medial knee displacement	1.0 (0.0-2.0)	1.0 (0.0-2.0)	0.59-0.93	0.60-0.95

Typical Error 90% CL~×/÷1.47, lowest ICC 90% CL~±0.47, highest ICC 90% CL~±0.02

The Pearson correlation coefficients suggested moderate to very large correlations between the peak ankle, knee and hip angles recorded during the functional tests and those recorded during jogging (r = 0.53 to 0.93; Table 3.5). The strongest correlations (r ≥ 0.70) for three or more functional tests existed for ankle eversion/inversion, knee abduction/adduction, hip abduction/adduction and hip internal/external rotation. The confidence limits for the majority of ankle, knee and hip correlations indicated the true correlations were very likely to be at least moderate (≥ 0.3) and likely to be large (≥ 0.5). The correlation for peak pelvic tilt was also high (r = 0.60 to 0.72), while trunk angles generally showed the poorest correlations (r = -0.15 to 0.53).

Table 3.5:   Correlations between peak joint angles during the functional tests and jogging expressed as Pearson correlation coefficients

		SKB SKB	Single Leg	Lunge	Hop Lunge	Stepdown
Ankle	Eversion	0.77	0.76	0.60	0.79	0.60
	Inversion	0.74	0.83	0.72	0.65	0.78
Knee	Abduction	0.70	0.70	0.79	0.66	0.76
	Adduction	0.84	0.75	0.69	0.61	0.78
Hip	Adduction	0.80	0.73	0.75	0.71	0.65
	Abduction	0.82	0.70	0.70	0.72	0.53
	Internal rotation	0.90	0.89	0.87	0.85	0.84
	External rotation	0.86	0.89	0.91	0.90	0.93
Pelvis	Lateral tilt	0.71	0.60	0.65	0.64	0.72
	Rotation	0.59	0.62	0.68	0.23	0.66
Trunk	Lateral flexion	0.09	0.10	0.18	-0.15	0.27
	Rotation	0.28	0.40	0.19	0.23	0.53

Pearson correlation coefficient 0.1 90% CL~±0.33, 0.6 90% CL~±0.22, 0.9 90% CL~±0.7
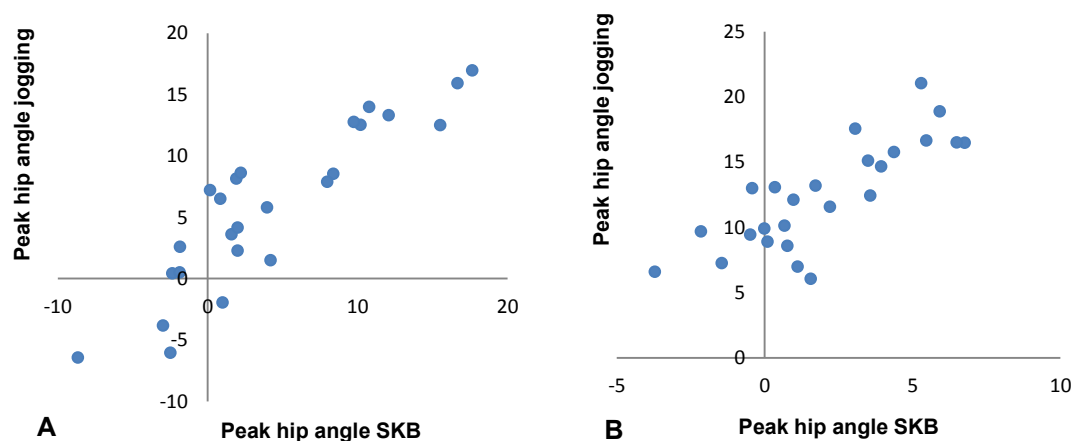
Figure 3.2:   Example scatter plots showing the association between SKB and jogging. Plot **A** shows peak transverse plane angles (-ve = external rotation), plot **B** shows peak frontal plane angles (-ve = abduction).

Discussion

The first purpose of this study was to investigate the within-day and between-day reliability of 3D joint kinematics during five lower extremity functional screening tests commonly used by physiotherapists in clinical practise. The findings indicated the within-day reliability of all kinematics was excellent and the between-day reliability of the majority of kinematics was good to excellent. To our knowledge this is the first study to report the 3D kinematic reliability of the trunk, pelvis, hip, knee and ankle during these specific functional tests. Previous studies have reported the reliability of some of these kinematics in similar but not identical tests such as stair descent (Bolgla, et al., 2008), drop jump (Ford, et al., 2007) and SLS (Zeller, et al., 2003). While these are obviously also useful assessment tests the instructions, test protocols and kinematics involved are subtly different to most of the SKB movement tests investigated in the current study (step-down test excepted).

Within-day reliability was better than between-day for all tests which supports the findings of previous studies investigating the reliability of 3D joint kinematics during gait and other lower extremity functional tests (Ford, et al., 2007; Kadaba, et al., 1989). Ford et al., (2007) reported on the reliability of peak lower extremity 3D kinematics when landing from a drop vertical jump in a group of 11 school soccer players. They concluded the reliability of the majority of kinematics was excellent to good (within-day ICC = 0.90, 95%CI 0.86-0.95, between-day ICC = 0.77, 95%CI 0.72-0.82). This is similar to the reliability reported for kinematics during normal adult gait (Kadaba, et al., 1989) and similar to the reliability seen in the current study. The typical errors reported by Ford et al., (2007) are also in accordance with those reported in the current study (within-day 0.9º to 3.2º, between-day 1.3º to 5.5º). Other studies have reported comparable reliability during other lower extremity functional tests. Earl et al., (2007) reported reliable between-day peak 3D joint kinematics (ICC ≥ 0.84) during a drop vertical jump and single-leg step-down, in a group of males and females in their early 20's, for variables including rearfoot eversion, knee flexion/abduction/internal rotation and hip adduction/internal rotation. Kernozek et al., (2005) reported good between-day reliability (ICC = 0.79 to 0.93) for a drop landing in a group of male and female recreational athletes (variables assessed were peak ankle plantarflexion/pronation, knee flexion/valgus and hip flexion/adduction). Bolgla et al., (2008) reported

similar reliability for mean hip joint kinematics (internal rotation and adduction) during the stance phase of a stair descent task (ICC = 0.75 to 0.88) in a group of females in their early 20's.

Of importance to clinical use is the excellent between-day reliability of peak hip kinematics in all planes in the current study. The role of hip motion in the frontal and transverse planes has been a focus of many recent studies investigating lower extremity alignment and risk of injury (Cowan, Crossley, & Bennell, 2008; Reiman, et al., 2009). Also of clinical relevance is our reporting of trunk and pelvis reliability which is not commonly reported in previous studies. The within-day reliability was similar to other segments as was the between-day reliability of the pelvis. However the between-day reliability of the trunk was somewhat less reliable. Additionally, except for the step-down test, knee flexion in the current study ranged from 65° to 68° This is in contrast to maximum angles of up to 96° (Earl, et al., 2007), 95° (Zeller, et al., 2003), and 89° (Kernozek, et al., 2005) reported in previous studies using different lower extremity functional tests. The lesser knee flexion in the current study may make the SKB tests more appropriate for clinical screening of trunk and lower extremity alignment during walking and jogging gait (we noted that the average maximum knee flexion during the stance phase of jogging in our participants was 45 ±6°). Previous authors investigating single leg squat have used additional monitoring to limit knee flexion to 45° (Levinger, et al., 2007) and 60° (Claiborne, et al., 2006), similar to the angles we recorded. Additionally, Willson and Davis (2008a) determined hip and knee kinematics of interest to be those at 45° knee flexion during a single leg squat. Maximum hip and knee flexion showed excellent reliability (ICC = 0.86 to 0.98) demonstrating the ability of participants to produce a consistent range of sagittal plane motion. This was achieved with very simple instructions to the participants and without the need for complicated and time consuming monitoring. Zeller et al., (2003) highlighted the need for controlling the depth of a single leg squat when assessing lower extremity kinematics and this has been noted as a limitation in other studies (Dwyer, Boudreau, Mattacola, Uhl, & Lattermann, 2010). In terms of test validity this is important as the maximum frontal and transverse plane deviations need to be assessed across the same range of sagittal plane motion on any given occasion for tests to be comparable. Increases in frontal plane motion, with increases in sagittal plane motion, have been reported in previous studies of kinematics during lower extremity functional tests (Kernozek, et al., 2005; Russell, Palmieri, Zinder, & Ingersoll, 2006; Zeller, et al., 2003).

As well as joint angles we also investigated the reliability of MKD (based on patella position) as this is commonly used by physiotherapists as a means of assessing dynamic lower extremity alignment. The excellent within-day reliability (ICC = 0.94 to 0.98) was again better than between-day (ICC = 0.62 to 0.94). However the between-day reliability was likely to be higher than reported as we failed to adequately standardise the medial position of the patella at the start of each test, between days. Thus we suggest that maximum MKD is likely to show acceptable reliability both within- and between-days. What remains unclear is whether or not the position of the patella is a good indicator of lower extremity 3D joint kinematics and this is a question that requires further research. It has been suggested that patella position in the frontal plane is a clinical indicator of femoral adduction and internal rotation (Bell, et al., 2008), but there appears to be little evidence to support this claim. Willson and Davis (2008b) recently described the 2-D frontal plane projection angle (FPPA) which uses an anterior knee marker at the mid-point of the femoral condyles (marker placed on the patella) to assess dynamic lower extremity alignment. They concluded that this angle was moderately associated with 3-D hip adduction and knee external rotation during single-leg squats, running and jumping.

There does not appear to be any consistent difference in the reliability between the various SKB tests and in fact the narrow range of typical errors suggests similar reliability. Intuitively we anticipated that the SKB (double leg stance throughout) may have been more reliable as it was the simplest test to perform and provided the least challenge to stability. This was however not the case and the within- and between-day typical errors reported for all kinematics, across all tests, were generally small (mostly between 1 and 4°). This gives an indication to physiotherapists as to the absolute variation they can expect for any given client between consecutive tests. However it must be recognised that in visual observation clinicians cannot decompose joint motion into its 3D components. Thus any variation observed visually will be a composite (addition) of the typical errors reported in the individual planes, further confounding visual assessment. Even so interpretation of the ICC's alongside these typical errors provides evidence for physiotherapists that what they are observing on any given occasion is repeatable and thus representative of a client's kinematics. As has been noted by previous authors this is better than interpretation of the ICC's alone due its sensitivity to the heterogeneity of values between participants (Hopkins, 2000). One caution here is that we repeated

the tests over a one to two day time period and this may not represent the typical error over longer timeframes.

Multiple factors are suggested to contribute to variability in repeat 3D motion analysis testing including errors due to marker placement. In contrast to previous studies we attempted to eliminate marker placement as a source of variation by marking the skin. This has been suggested by others as a way to improve the accuracy of marker reapplication (Ford, et al., 2007). We also used standardised marker placement by a single investigator. The variability observed between tests is thus likely to mostly represent the altered movement performance of the individual participants. This was the major source of variation of interest in the current study as it is the most clinically relevant when considering the use of these functional tests. When considering clinical use of the tests it was encouraging to see the reliability of kinematics in the frontal and transverse planes as these are the movements most commonly linked to risk of lower extremity injury. We did investigate the influence of an increased number of trials on the reliability of the functional tests. This additional processing of errors suggested that the use of three trials (as is likely to occur in common clinical practice) provides similar reliability to the use of ten trials. The performance of extra trials does not appear warranted given the time that would be involved and the very small reductions in typical errors predicted.

A secondary purpose of this study was to investigate the association between the peak 3D joint kinematics during the functional tests and those occurring during jogging. The results show that for many of the variables of interest there is a moderate to strong association suggesting that participants with higher peak angles in the functional tests also had higher peak angles in jogging. A caution here is that this should not be interpreted as the absolute angles agreeing as can be seen from the scatter plots in Figure 3.2. This association does however provide some preliminary support for the use of these tests as a screening tool and specifically for screening peak lower extremity kinematics during the loading period of the stance phase in jogging gait. There are obvious clinical advantages involving space and equipment which make the use of SKB tests more feasible than direct assessment of jogging gait. As the velocity of the movement during SKB is also slower than with jogging it is likely that physiotherapists will be able to visually rate movement with greater reliability and validity. This association between SKB and jogging appears strongest for hip kinematics. As noted earlier there

has been increasing interest recently in the role of the hip in knee dysfunction (Reiman, et al., 2009) We further caution here however that we are not suggesting the SKB tests can take the place of jogging gait assessment. We have not performed a comparison of kinematics throughout (or at specific time points during) the SKB tests and jogging gait cycle which would be required before this could be contemplated.

A further limitation of this study is the uninjured population that we investigated. The reliability of these lower extremity functional tests in a population with lower extremity injury, such as patellofemoral syndrome, needs further investigation. However the current results provide useful information for the use of the tests when attempting to predict future injury in a screening situation. Another limitation is the small sample size used to assess the reliability of the functional tests. This obviously leads to greater uncertainty in the true magnitude of the reliability in the population. Precision for a correlation is thought to be adequate when the uncertainty in the estimate (represented by its confidence interval) does not span more than two qualitative magnitude thresholds (Hopkins & Manly, 1989). Thus we appear to have adequate precision for within-day ICC's and Pearson correlations greater than 0.8 and for between-day ICC's of greater than 0.9. The lower correlations still provide useful information but they must be interpreted with more caution. A final limitation of the study is our lack of control over jogging velocity and thus we cannot state that participants were running at a constant velocity. It should also be noted that peak 3D transverse and frontal plane hip and knee angles may occur in the second half of stance phase of jogging and our analysis only included the first half.

Conclusion

In healthy participants the reliability of peak 3D kinematics during the descent phase of these lower extremity functional tests is acceptable for the majority of kinematic variables of interest to physiotherapists. There is a moderate to strong association between these peak kinematics and those recorded during jogging. Based on these results SKB lower extremity functional movement tests should be useful in helping physiotherapists make clinical decisions regarding trunk and lower extremity dynamic alignment and risk of injury.

CHAPTER 4: PHYSIOTHERAPIST AGREEMENT WHEN VISUALLY RATING MOVEMENT QUALITY DURING LOWER EXTREMITY FUNCTIONAL SCREENING TESTS

Overview

Objectives: To investigate physiotherapist agreement in rating movement quality during lower extremity functional tests using two visual rating methods and physiotherapists with differing clinical experience. *Design:* Clinical measurement. *Participants:* Six healthy individuals were rated by 44 physiotherapists. These raters were in three groups (*inexperienced, novice, experienced*). *Main measures:* Video recordings of all six individuals performing four lower extremity functional tests were visually rated (*dichotomous* or *ordinal* scale) using two rating methods (*overall* or *segment*) on two occasions separated by 3-4 weeks. Intra and inter-rater agreement for physiotherapists was determined using overall percentage agreement (OPA) and the first order agreement coefficient (AC1). *Results:* Intra-rater agreement for overall and segment methods ranged from slight to almost perfect (OPA: 29% to 96%, AC1: 0.01 to 0.96). AC1 agreement was better in the *experienced* group (84 to 99% likelihood) and for *dichotomous* rating (97 to 100% likelihood). Inter-rater agreement ranged from fair to good (OPA: 45% to 79%; AC1: 0.22 to 0.71). AC1 agreement was not influenced by clinical experience but was again better using *dichotomous* rating. *Conclusions:* Physiotherapists' visual rating of movement quality during lower extremity functional tests resulted in slight to almost perfect intra-rater agreement and fair to good inter-rater agreement. Agreement improved with increased level of clinical experience and use of dichotomous rating.

Background

Visual assessment of movement quality is common in physiotherapy and is an important component of a therapists' decision making process regarding intervention (Bernhardt, et al., 1998). There is evidence these visual assessments are accurate and reliable for rating upper limb movement and gait in subjects with neurological impairments (Bernhardt, et al., 1998; Lord, et al., 1998). However,

accuracy and reliability of visual rating of lower extremity movement quality in subjects with current or potential musculoskeletal disorders is less well understood. Visual rating of this population during various tasks appears common clinically, often focused on frontal plane control of the pelvis, hip and knee as this has frequently been linked to injury (Mascal, Landel, & Powers, 2003; Powers, 2003; Souza & Powers, 2009; Willson & Davis, 2009). Sahrmann (2002) referred to the term movement quality, emphasising that in the assessment and treatment of musculoskeletal disorders, movement quality should be emphasised rather than quantity (number of repetitions or time taken).

A recent focus on the importance of lower extremity dynamic alignment in many lower extremity injuries, particularly patellofemoral dysfunction (Levinger, et al., 2007; Souza & Powers, 2009; Willson & Davis, 2008b) further emphasises the need for this to be assessed. Poor dynamic alignment is a combination of poor frontal and/or transverse plane control of the trunk, pelvis, hip, knee and foot (Earl, et al., 2005; Powers, 2003; Sahrmann, 2002; Willson & Davis, 2009). Excessive medial displacement of the knee in the frontal plane has been termed dynamic knee valgus (Bell, et al., 2008). Clinical observation of the position of the patella relative to a line extending vertically from the 1st or 2nd toes is a common measure of this frontal plane knee control (Bell, et al., 2008; Hirth & Padua, 2007; Mottram & Comerford, 2008). Recently Powers (2010) presented a biomechanical argument suggesting that trunk, pelvis and hip control (predominantly in the frontal and transverse planes) should be addressed in the design of all knee rehabilitation programs. Improvements in frontal and transverse plane pelvis and hip control have accompanied a reduction in patellofemoral pain (Mascal, et al., 2003) and lateral trunk control has been a predictor of knee injury in female college athletes (Zazulak, Hewett, Reeves, Goldberg, & Cholewicki, 2007). A prospective study of physical education students demonstrated that increased pronation was a risk factor for the development of exercise related lower limb pain (Willems, et al., 2006).

Visual observation of functional tests is currently the most common method of assessing dynamic alignment in the clinic. Furthermore functional tests are a simple way of simulating a painful activity while allowing therapists to qualitatively assess movement quality through visual observation (Bolgla & Keskula, 1997; Clark, 2001). Functional tests are also used as a screening tool to help predict an individual's risk of future injury as they are aimed at simulating a movement pattern common to many

athletic activities including landing, squatting, lunging, and the stance phase of running (Kibler, et al., 2002; Noyes, et al., 2005; Reid, et al., 2003; Zeller, et al., 2003). Mottram and Comerford (2008) suggested that screening assessments need to have a greater focus on functional tasks that seek to evaluate multiple muscle interactions acting on multiple joints. Many assessments described by Mottram and Comerford (2008) appear to be based on visual observation. Therapists rely on visual analysis as they often do not have access to the equipment or time required for complex biomechanical analysis. As therapists are using visual ratings to make clinical decisions, the reliability of these ratings needs to be considered.

The small knee bend (SKB) lower extremity functional movement (also known as a partial squat) and its variations are commonly used by physiotherapists and sports physicians to assess dynamic lower extremity alignment. An association between peak lower extremity kinematics during SKB and jogging has recently been reported (Whatman, et al., 2011a). Both dual limb SKB (Sahrmann, 2002) and single leg variations of the SKB test (Reid, et al., 2003; Sahrmann, 1998) have been used as part of a lower quarter examination. The single leg SKB of Sahrmann (1998) is similar to the unilateral squat described and depicted by Chmielewski et al. (2007) and a single leg SKB has also been described as a screening test by Mottram and Comerford (2008). When indicated in clinical practice, further evaluation of dynamic lower extremity alignment commonly involves a lunge (Crossley, et al., 2006) and hop lunge (Cook, 2006). Similar movements used for testing/screening reported in the literature include single leg squats, single leg step downs (Chmielewski, et al., 2007; Willson, et al., 2006), lunges (Thijs, et al., 2007) and drop landings (Noyes, et al., 2005).

Although functional tests are commonly used, the reliability of visually rating movement quality, especially of multiple body segments, has not been well defined. Given clients are likely to see a single physiotherapist, intra-rater agreement is most important, however inter-rater agreement also needs investigation. Most recently Ekegren, Miller, Celebrini, Eng and Macintyre (2009) reported the reliability and validity of simple observational screening to detect risk of anterior cruciate ligament injury. Three experienced physiotherapists showed high intra-rater agreement (Kappa(κ)=0.75 to 0.80), but lacked sensitivity when rating uninjured female soccer players as high or low risk based on patella position during a drop-jump task. Inter-rater agreement was also high (κ=0.77 and 0.80) and

these authors noted that their simple (solely based on observation of patella position) dichotomous rating method was key to the high level of agreement reported. In contrast Chmielewski et al. (2007) reported much lower intra-rater and inter-rater reliability in their study (κ=0.01 to 0.68) investigating visual rating during a unilateral squat and lateral step-down task. These authors used an overall and a segmental rating method (rating multiple body segments) that more closely resembled clinical practice. This approach is supported by Mottram and Comerford (2008) who recommended rating the trunk, pelvis, knee and foot separately when assessing the SKB. Although this segmental approach has been recommended when screening for risk of injury (Mottram & Comerford, 2008), overall ratings classifying movement deviations as minor, moderate or marked are also common in clinical practice (Reid, et al., 2003). Chmielewski et al. (2007) concluded that further research is needed to develop visual analysis of movement, suggesting future studies include ratings of individual body segments and use three point or dichotomous scales to indicate the severity of movement deviations. We are not aware of any studies that have investigated visual ratings of multiple body segments across multiple common clinical tests including the lunge or hop lunge. Therefore further research is required to establish the reliability of visually rating movement quality during common lower extremity functional tests using both overall and segmental rating methods. Both methods have been recommended in the literature and are common in current physiotherapy clinical practice. The use of dichotomous and ordinal scales for rating movement also requires further investigation.

From the studies cited above the reliability of visual rating appears to depend on the movement being rated and the method of rating. Clinical experience may also affect the reliability of visual ratings. Previous studies have only used a small number of similarly experienced physiotherapists and the results may not be generalisable to all physiotherapists. More experience has been suggested to improve the reliability of visual gait assessment in patients with musculoskeletal disorders (Brunnekreef, van Uden, van Moorsel, & Kooloos, 2005) and the high agreement in the study of Ekegren et al. (2009) only involved experienced physiotherapists. However another study comparing undergraduate students to experienced physiotherapists found no difference in reliability with visual assessment of static foot position (Somers, Hanson, Kedzierski, Nestor, & Quinlivan, 1997). Furthermore we are not aware of any studies that have attempted to quantify the effect of experience

on the level of agreement. Thus the effect of physiotherapist experience on the reliability of visually rating lower extremity functional tests warrants further investigation.

Purpose

To investigate physiotherapist agreement in rating movement quality during commonly used lower extremity functional tests, using two visual rating methods which capture visual ratings in a manner that is consistent with current clinical practice and physiotherapists with differing clinical experience.

Methods

Participants

Video recordings of six (four female) healthy individuals (mean ±SD, age = 22 ±3 yrs, height = 170 ±9 cm, weight = 67 ±15 kg) performing four lower extremity functional tests (Small knee bend, Single leg small knee bend, Lunge, Hop lunge) were visually assessed using two rating methods (Overall, Segment). These participants were selected from a group of twenty five participants (university students) with no musculoskeletal problems who volunteered to be videoed for a larger study. Physiotherapists in three groups of increasing level of clinical experience (Inexperienced, Novice, Experienced) performed the assessments on two separate occasions 3-4 weeks apart. A period of two weeks has been suggested as the minimum to avoid raters remembering initial ratings (Ekegren, et al., 2009). The study was approved by the University Ethics Committee and all participants and physiotherapists received verbal and written information about the study and gave written informed consent. The number of participants rated was primarily a decision based on how much time we could reasonably expect physiotherapists in practice to give to the study on two occasions (when piloting our visual rating scale we estimated it would take approximately 45 to 60 minutes to view and rate six individuals).

Physiotherapists (raters)

Forty four physiotherapists, comprising three groups categorized based on level of training and years of experience (Inexperienced: 14 fourth year undergraduate physiotherapy students; Novice: 18 musculoskeletal postgraduate physiotherapy students with (mean ±SD) 4 ±2 years clinical experience; Experienced: 12 experienced physiotherapists with 15 ±4 years clinical experience and all with postgraduate qualifications in musculoskeletal physiotherapy), participated in the study. The students were recruited based on those who were available at appropriate times. All those who indicated interest were invited to participate. The Experienced physiotherapists were all required to have a postgraduate musculoskeletal qualification and they were individually invited to participate. Seventy five percent of the physiotherapists (12 Inexperienced, 11 Novice and 10 Experienced) were available for the repeat assessments.

Visual assessment procedure

For all four lower extremity functional tests, physiotherapists visually assessed overall movement pattern quality using the Overall rating method, as well as segment movement quality using the Segment rating method for each of the trunk, pelvis, knee and foot. Segment ratings were based on a judgement as to whether participants maintained an acceptable segment position throughout each test (see Figure 4.1). Descriptions of acceptable segment movement were based on those previously reported by Chmielewski et al. (2007), common clinical practice and discussions with three expert physiotherapists. Included in the Segment method was also a rating of "oscillation" (repetitive movement of any of the lower extremity segments to and from neutral). Oscillation is considered a movement dysfunction (greater magnitudes and frequencies indicating greater dysfunction) (Chmielewski, et al., 2007). Increased oscillation of the lower extremity has been linked to risk of knee injury in female athletes (Ford, et al., 2006). Overall ratings were made separately from the segment ratings and required the categorisation of movement quality as acceptable, minor, moderate or marked dysfunction. No specific guidelines were given for scoring the severity of overall movement dysfunction. All ratings were recorded on a standardised rating sheet specifically designed for this study (see Figure 4.1). For scoring segment movement quality the rating sheet included a yes/no rating to indicate if a segment movement was acceptable and a three point scale (minor, moderate,

marked) to score the severity of any unacceptable ratings. All aspects of the rating sheet were designed in consultation with three expert physiotherapists (all PhD or Masters qualified and previously part of Olympic medical teams) who piloted its use and had input into the final design. The Inexperienced and Novice groups of physiotherapists performed all ratings in a classroom situation where the video clips were played on a large screen via a projector. Both groups were instructed not to compare ratings with fellow students. The Experienced group of physiotherapists received a CD ROM containing the video clips to be rated. In both situations an attempt at randomisation was made by using the "shuffle" function in Windows™ Media Player. Video clips were produced using Windows™ Movie Maker and title screens were used to identify the participant number and test. Standardised instructions on how to view the video clips and use of the visual rating sheet were given to all physiotherapists prior to them performing visual ratings. Video clips were set up such that each functional test, performed by each individual on the CD, was looped four times (each repetition was thus identical). Based on pilot testing using the rating sheet, video clips were paused after two repetitions (to allow for ratings of the trunk and pelvis) and then again after the fourth repetition to allow for the remainder of the segment ratings and an overall rating. Pausing of the video clips during the performance of a functional test was not permitted. All physiotherapists were allowed to repeat video clips as many times as they felt necessary to be confident with their ratings. Any repeats of the videos were recorded during the classroom sessions and the experienced physiotherapists recorded any repeats on their rating sheets. This allowed us to calculate the total viewing time (time actually watching the video clips) spent by each of the experienced physiotherapists (combined length of the video clips plus recorded repeats). A small number of repeats were recorded by all groups and the viewing time spent by the experienced group (mean ±SD, 12 ±1 mins) was very similar to the time of the other two groups (Novice = 14 mins; Inexperienced = 15 mins).

Lower extremity functional tests - visual rating sheet		
Please rate movement quality by filling in the sheet below for each of the **four tests** you view on the video clips. Please circle **N** (No) OR **Y** (Yes) for oscillation and each segment below (trunk, pelvis, knee, foot). For any **YES** rating please also grade by circling **1** (minor), **2** (moderate) or **3** (marked). Please also rate **overall movement quality** for each test.		
Trunk:	Moves out of neutral in frontal or transverse plane	[N] [Y; 1,2,3]
Pelvis 1:	Moves out of neutral in the frontal or transverse plane	[N] [Y; 1,2,3]
Pelvis 2:	Moves away from the midline	[N] [Y; 1,2,3]
Knee:	Patella moves out of line with 2nd toe	[N] [Y; 1,2,3]
Foot:	Moves into excessive pronation	[N] [Y; 1,2,3]
Oscillation:	Observable oscillation (movement to and from neutral)	[N] [Y; 1,2,3]
Overall movement quality:	Acceptable movement pattern	0
	Minor movement dysfunction	1
	Moderate movement dysfunction	2
	Marked movement dysfunction	3

Figure 4.1:   Visual rating sheet used in the study allowing both Segment and Overall ratings.

Four lower extremity functional movement tests

The four lower extremity functional tests shown on the CD ROM were chosen due to their common usage by physiotherapists in musculoskeletal practice and their ease of administration. It was hoped that presenting four variations of the Small knee bend would increase the variety of movement patterns to be rated by the physiotherapists. The six individuals included on the CD ROM were chosen by the lead researcher from a collection of 25 individuals who had been recorded as part of another study. The six individuals were all right leg dominant (based on preferred kicking leg) and

were chosen in an attempt to include movement patterns across the range of the rating scale. All individuals performed all four movement tests (see Table 4.1) barefoot, with their arms held relaxed by their side and wore a tight fitting sleeve-less shirt (rolled up to expose their lower trunk/upper pelvis) and a pair of tight fitting shorts. All individuals on the video were given standardized verbal instructions prior to each test and the researcher demonstrated each test in a standardized manner. Practice for all tests was allowed until the researcher was happy the test was performed appropriately and consistently. All individuals performed three repetitions of each test. For the purposes of another study the individuals were instrumented with 15 retro-reflective markers (19 mm diameter) secured to specific anatomical locations (bilateral acromion processes, bilateral ASIS's, sacrum, bilateral iliac crests, medial and lateral femoral epicondyles, mid-patella, medial and lateral malleoli, head of 5th metatarsal, head of 1st metatarsal, posterior calcaneus). Additional detail on these particular functional tests and their reliability has been published previously (Whatman, et al., 2011a).

Table 4.1:   Description of the four lower extremity functional movement tests.

Functional test	Test description
Small knee bend (SKB)	Starting from a standing position, individuals performed a partial squat (hip and knee flexion) with the trunk maintained in an upright position. Individuals were instructed to continue the SKB until they reached maximum dorsiflexion without lifting their heels and then return to upright standing.
Single leg small knee bend (dominant leg*)	Standing on the dominant leg only, with the contralateral hip in neutral and contralateral knee flexed to approximately 80°, individuals performed a SKB as described above.
Lunge (dominant leg)	From a standing position individuals were instructed to lunge forward (leading with their dominant leg) a distance of approximately one and a half times the length of their normal gait stride. As they moved into single leg stance (on the dominant leg, with the contralateral leg off the ground) they flexed the hip and knee while maintaining an upright trunk. Individuals were instructed to continue the lunge until reaching maximum dorsiflexion of the stance leg without lifting their heel.
Hop lunge (dominant leg)	From a standing position individuals were instructed to jump forward a distance of approximately 1.0 m and on landing on the dominant leg to flex the hip and knee. Individuals were instructed to continue the lunge until reaching maximum dorsiflexion of the dominant leg without lifting their heel.

*Based on preferred kicking leg.

The four functional tests were recorded from an anterior view on digital video (Panasonic, USA) sampling at a rate of 60 Hz. The video camera was positioned on a tripod in front of individuals and perpendicular to the frontal plane, at a height of 0.86 m and a distance of 3.7 m. The zoom function of the camera was used to allow the frame of view to capture the individual from the shoulders down.

Statistical analyses

The level of intra-rater and inter-rater agreement (based on initial ratings only) for physiotherapists was determined using overall percentage agreement and the first order agreement coefficient (AC1). For both the Segment and Overall rating methods the agreement reported is based on each physiotherapist's ratings across all four functional movement tests (i.e. for each segment agreement is based on 24 ratings), and it has been analysed in two ways: (i) Based on dichotomous classification of ratings (yes/no for the segment method and acceptable/not acceptable for the overall method) (Dichotomous rating); (ii) Based on the four point ordinal classification of ratings (no/acceptable, minor, moderate and marked) (Ordinal rating).

Overall percentage of agreement was calculated as the number of observed agreements divided by the maximum number of possible agreements. The first order agreement coefficient (AC1) was calculated using a SAS macro published by Blood and Spratt (2007). Table 4.2 gives further detail as to how this statistic is calculated for two raters and the formula for the general case involving multiple raters and categories has been described (Blood & Spratt, 2007). Using the data format shown in Table 4.2, where ratings were made by two raters on a dichotomous scale (1 or 2), Chan (2003) described the calculation of the AC1 statistic as; $AC1 = (p-\emptyset)/(1-\emptyset)$ where $p = (A + D)/N$ and $\emptyset = 2q(1-q)$, $q = (A1 + B1)/2N$. Blood and Spratt (2007) described the formula for the more general situation of more than two raters and a rating scale with greater than two categories (see Figure 4.2). The AC1 adjusts the overall percentage agreement for chance agreement giving an estimate of the true agreement. The AC1 is a statistic proposed by Gwet (2008a) which is becoming increasingly popular for assessing agreement in ratings where high trait prevalence and/or rater bias is thought likely. Under these conditions the AC1 is suggested to be more stable and consistent with percentage agreement than the better known kappa statistic. Kappa is sensitive to prevalence and bias which can lead to the paradox of kappa, high percentage agreement but low kappa (Feinstein & Cicchetti, 1990) such that the kappa does not reflect the true agreement between raters. We anticipated there would be high prevalence of 'No' and 'Yes (1)' ratings given the healthy uninjured nature of our participants and thus this paradox of kappa would be a problem. Gwet (2008a) suggested that kappa only gives a reasonable measure of agreement when prevalence is close to 50% and its ability to reflect agreement diminishes considerably as prevalence gets closer to 0 or 100%. If prevalence is around

50% both the kappa and AC1 statistics perform alike. The AC1 was used in all inferences as to the magnitude of agreement and any effects of rating method or level of experience. Gwet (2001) suggested the AC1 can be interpreted in a similar manner to the kappa coefficient, based on a scale proposed by Landis (1977): 0.01-0.20=slight; 0.21-0.40=fair; 0.41-0.60=moderate; 0.61-0.80=good; 0.81-1.0=almost perfect.

Excel spreadsheets (Hopkins, 2007) were used to compare the agreement of the three physiotherapy groups (Inexperienced, Novice and Experienced) and to compare dichotomous versus ordinal ratings (further details of these statistical methods are given in the appendix). As suggested by Hopkins, Marshall, Batterham and Hanin (2009) all conclusions were based on magnitude based inferences rather than null-hypothesis testing. Rather than the use of p values we have used inferences based on the uncertainty in the magnitude of the differences between groups through the use of confidence intervals. Inferences were based on where the confidence interval lay in relation to threshold values for substantial differences and were qualified by probabilities using a scale published by Hopkins et al. (2009). Readers unfamiliar with this type of analysis are referred to Hopkins et al. (2009) for a more detailed explanation.

Table 4.2:   Data format used for the calculation of the AC1 statistic.

Rater B	Rater A		
	1	**2**	**Total**
1	A	B	**B1** = A+B
2	C	D	**B2** = C+D
Total	**A1** = A+C	**A2** = B+D	N

Results

Ratings for all functional tests were somewhat clustered, irrespective of the rating method used. Most of the Overall ratings (71%) and Segment ratings (75%) were in the acceptable and minor dysfunction categories.

Intra-rater agreement for the 33 physiotherapists who made repeat ratings ranged from slight to almost perfect, for overall movement pattern quality as well as segment movement quality (OPA: 29% to 96%, AC1: 0.01 to 0.96) (see Table 4.3). Experienced physiotherapists had the highest level of intra-rater agreement and also appeared less variable in the overall rating, as suggested by the narrower AC1 range (0.72 to 0.85). Irrespective of the rating method used (Overall or Segment), the intra-rater agreement of Experienced (AC1 range 0.47 to 0.80) was likely/very likely (84%-99%) to be better than agreement in the other two groups (see Table 4.5). Based on Ordinal ratings Inexperienced possibly (74%-84% likelihood) reached better intra-rater agreement than Novice. All physiotherapists almost certainly (93%-100% likelihood, see Table 4.7) showed better intra-rater agreement based on Dichotomous ratings (agreement moderate to good for most ratings) than on Ordinal ratings (agreement fair to moderate). There appeared to be little consistent difference in the level of intra-rater agreement between the Segment and Overall rating methods.

Inter-rater agreement was generally not as good as intra-rater agreement. Inter-rater agreement for each group and for all physiotherapists combined ranged from fair to good for overall movement pattern quality as well as segment movement quality (OPA: 45% to 79%; AC1: 0.22 to 0.71) (see

Table 4.4). Experience was much less likely to improve inter-rater agreement compared to intra-rater agreement. However Experienced and Novice possibly (74% and 76% likelihood) reached better agreement than Inexperienced based on Dichotomous ratings (see Table 4.6). In the same way as intra-rater agreement, inter-rater agreement did not seem to be consistently influenced by the rating method (Overall versus Segment) but again was very likely to be better on Dichotomous ratings (76%-100% likelihood, see Table 4.7).

Table 4.3:   Mean intra-rater agreement for the Segment and Overall rating methods.

	Percent Agreement (range)		AC1 (range)	
	Dichotomous	**Ordinal**	**Dichotomous**	**Ordinal**
Segment method				
Trunk				
Experienced	79 (75-80)	67 (58-80)	0.63 (0.22-0.85)	0.59 (0.35-0.79)
Novice	75 (57-83)	58 (42-79)	0.54 (0.31-0.68)	0.47 (0.26-0.76)
Inexperienced	75 (50-83)	62 (38-80)	0.52 (0.03-0.73)	0.53 (0.24-0.65)
Pelvis 1				
Experienced	82 (75-85)	68 (62-83)	0.67 (0.33-0.81)	0.59 (0.37-0.79)
Novice	79 (50-92)	54 (33-71)	0.63 (0.10-0.87)	0.42 (0.12-0.63)
Inexperienced	77 (63-96)	63 (38-84)	0.57 (0.28-0.92)	0.53 (0.19-0.69)
Pelvis 2				
Experienced	83 (78-86)	64 (58-86)	0.69 (0.33-0.81)	0.54 (0.21-0.79)
Novice	72 (42-88)	52 (33-71)	0.48 (0.10-0.80)	0.40 (0.06-0.62)
Inexperienced	78 (50-92)	63 (29-75)	0.59 (0.08-0.92)	0.54 (0.22-0.74)
Knee				
Experienced	84 (70-93)	59 (48-70)	0.73 (0.26-0.96)	0.47 (0.26-0.74)
Novice	84 (75-96)	55 (38-71)	0.75 (0.57-0.95)	0.42 (0.19-0.63)
Inexperienced	80 (54-92)	57 (46-80)	0.64 (0.12-0.92)	0.44 (0.08-0.68)
Foot				
Experienced	74 (68-76)	61 (52-75)	0.54 (0.15-0.88)	0.53 (0.40-0.75)
Novice	66 (50-83)	52 (38-75)	0.39 (0.01-0.68)	0.41 (0.10-0.69)
Inexperienced	70 (46-88)	62 (46-80)	0.47 (0.01-0.86)	0.54 (0.15-0.68)
Oscillation				
Experienced	88 (77-94)	65 (58-75)	0.80 (0.57-0.96)	0.57 (0.40-0.75)
Novice	81 (67-88)	53 (38-67)	0.69 (0.37-0.85)	0.40 (0.19-0.59)
Inexperienced	81 (67-92)	61 (46-80)	0.68 (0.34-0.96)	0.51 (0.19-0.67)
Overall method				
Experienced	87 (76-93)	68 (60-75)	0.79 (0.72-0.85)	0.59 (0.43-0.74)
Novice	80 (63-92)	53 (38-79)	0.68 (0.38-0.87)	0.40 (0.16-0.74)
Inexperienced	82 (71-94)	64 (54-79)	0.72 (0.42-0.94)	0.54 (0.18-0.67)

AC1 = overall percentage agreement adjusted for chance agreement giving an estimate of the true agreement.

Table 4.4: Inter-rater agreement for the Segment and Overall rating methods.

	Percent Agreement		AC1 (95% CL)	
	Dichotomous	**Ordinal**	**Dichotomous**	**Ordinal**
Segment method				
Trunk				
Experienced	65	49	0.31 (0.19-0.43)	0.35 (0.27-0.44)
Novice	70	51	0.40 (0.31-0.50)	0.38 (0.29-0.47)
Inexperienced	70	53	0.39 (0.26-0.53)	0.41 (0.29-0.53)
All physiotherapists	69	51	0.37 (0.28-0.47)	0.39 (0.30-0.47)
Pelvis 1				
Experienced	71	48	0.44 (0.29-0.59)	0.33 (0.23-0.42)
Novice	72	46	0.49 (0.38-0.60)	0.29 (0.21-0.37)
Inexperienced	71	52	0.41 (0.27-0.56)	0.39 (0.26-0.52)
All physiotherapists	70	47	0.42 (0.31-0.54)	0.32 (0.23-0.41)
Pelvis 2				
Experienced	70	47	0.42 (0.27-0.56)	0.32 (0.23-0.41)
Novice	70	45	0.47 (0.36-0.58)	0.29 (0.23-0.35)
Inexperienced	75	54	0.50 (0.34-0.65)	0.42 (0.30-0.55)
All physiotherapists	70	48	0.43 (0.32-0.54)	0.32 (0.25-0.40)
Knee				
Experienced	76	48	0.60 (0.47-0.74)	0.33 (0.26-0.40)
Novice	81	50	0.71 (0.58-0.84)	0.35 (0.30-0.41)
Inexperienced	75	43	0.59 (0.47-0.71)	0.25 (0.15-0.35)
All physiotherapists	78	47	0.64 (0.53-0.76)	0.30 (0.24-0.36)
Foot				
Experienced	61	45	0.22 (0.10-0.34)	0.30 (0.24-0.37)
Novice	63	46	0.28 (0.16-0.39)	0.32 (0.25-0.39)
Inexperienced	60	49	0.24 (0.13-0.34)	0.38 (0.27-0.48)
All physiotherapists	62	47	0.23 (0.14-0.32)	0.33 (0.26-0.40)
Oscillation				
Experienced	79	50	0.68 (0.55-0.80)	0.37 (0.30-0.44)
Novice	80	48	0.71 (0.60-0.82)	0.34 (0.28-0.40)
Inexperienced	79	47	0.67 (0.55-0.79)	0.31 (0.25-0.37)
All physiotherapists	79	48	0.68 (0.57-0.79)	0.33 (0.28-0.38)
Overall method				
Experienced	76	49	0.60 (0.45-0.75)	0.35 (0.27-0.42)
Novice	77	50	0.58 (0.45-0.71)	0.38 (0.30-0.46)
Inexperienced	70	45	0.46 (0.34-0.58)	0.28 (0.21-0.36)
All physiotherapists	76	49	0.61 (0.48-0.73)	0.34 (0.28-0.40)

CL = confidence limits.

Table 4.5:   Differences in mean intra-rater AC1, (90% CL), percentage likelihood the difference is clinically meaningful (> than the smallest worthwhile difference, Cohen 0.2), [group with the higher agreement E=experienced, N=novice, I=inexperienced].

	Experienced v Novice	Experienced vs Inexperienced	Novice vs Inexperienced
Dichotomous ratings			
Segment method	0.12 (0.00-0.21), 89% [E]	0.12 (0.00-0.21), 89% [E]	0.00 (-0.10-0.08), 31% [N]
Overall method	0.09 (0.02-0.14), 95% [E]	0.06 (0.00-0.11), 87% [E]	0.03 (-0.07-0.11), 51% [I]
Ordinal ratings			
Segment method	0.14 (0.06-0.20), 99% [E]	0.07 (-0.01-0.14), 84% [E]	0.07 (-0.01-0.13), 84% [I]
Overall method	0.19 (0.06-0.30), 96% [E]	0.10 (-0.02-0.20), 84% [E]	0.09 (-0.05-0.20), 74% [I]

Table 4.6:   Difference in inter-rater AC1, (90% CL), percentage likelihood the difference is clinically meaningful (> Fisher 0.1), [group with the higher agreement E=experienced, N=novice, I=inexperienced].

	Experienced v Novice	Experienced v Inexperienced	Novice v Inexperienced
Dichotomous ratings			
Segment method	0.08 (0.02-0.13), 56% [N]	0.02 (-0.05-0.08), 7% [E]	0.10 (0.04-0.15), 76% [N]
Overall method	0.02 (-0.18-0.16), 33% [E]	0.14 (-0.04-0.27), 74% [E]	0.12 (-0.04-0.24), 69% [N]
Ordinal ratings			
Segment method	0.00 (-0.03- 0.03), 0%	0.00 (-0.03-0.05), 0%	0.00 (-0.03-0.05), 0%
Overall method	0.03 (-0.06-0.12), 16% [E]	0.07 (-0.02-0.15), 35% [E]	0.10, (0.01-0.19), 57% [N]

Table 4.7: Difference in agreement (AC1) between the Dichotomous ratings and the Ordinal ratings (90% CL), percentage likelihood the difference is clinically meaningful (intra > Cohen 0.2, inter > Fisher 0.1).

	Intra-rater	Inter-rater
Segment method		
Experienced	0.14 (0.05-0.21), 97%	0.14 (0.09-0.19), 96%
Novice	0.18 (0.10-0.25), 100%	0.21 (0.17-0.25), 100%
Inexperienced	0.09 (0.01-0.16), 93%	0.11 (0.05-0.16), 76%
Overall method		
Experienced	0.18 (0.11-0.30), 100%	0.25 (0.09-0.37), 96%
Novice	0.28 (0.17-0.37), 100%	0.20 (0.06-0.31), 93%
Inexperienced	0.24 (0.14-0.32), 100%	0.18 (0.06-0.29), 89%

Discussion

The purpose of this study was to investigate physiotherapist intra and inter-rater agreement when visually rating movement quality via video during four lower extremity functional tests. The influence of variations in the experience level of the physiotherapists (Inexperienced, Novice and Experienced), rating method (Segment method versus Overall method) and classification of ratings (Dichotomous ratings versus Ordinal ratings) were also examined. Where appropriate we presented the likelihood the magnitude of any differences in level of agreement were clinically important, which to our knowledge has not been reported in the past.

Mean intra-rater agreement was mostly moderate to good for all groups of physiotherapists. This was achieved without detailed instructions or training, using a rating method designed to reflect current clinical practice, suggesting this level of agreement is achievable in the clinic. Although difficult to compare, due to differences in design and analysis, the level of intra-rater agreement reached in this study appears similar or better than reported previously (Chmielewski, et al., 2007; Crossley, Bryant, Jing Zhang, & Cowan, 2006; Ekegren, et al., 2009). In a study involving uninjured participants of a similar age, Chmielewski et al. (2007) reported intra-rater percent agreement of 32% to 76% when visually rating movement quality during a unilateral squat and lateral step-down task. These authors used both a segment (trunk, pelvis, hip) and overall method of rating movement quality based on four criteria and gave raters training prior to the visual assessments. In a study appearing to use a similar overall method to rate movement quality as poor or good during a single leg squat, Crossley, Bryant,

Zhang and Cowan (2006) reported acceptable levels of reliability for experienced physiotherapists. Similarly, Ekegren et al. (2009), reported substantial intra-rater percent agreement (88 to 90%) for visual ratings of dynamic knee valgus, during a drop jump performed by healthy participants. These authors used a simple rating method based on patella position and also gave raters (three experienced physiotherapists) detailed instructions and training.

The level of inter-rater agreement achieved by the physiotherapists was fair to good. This was similar to the inter-rater percent agreement reported (32% to 82%) in the study by Chmielewski et al. (2007). Similarly another study by DiMattia, Livengood, Uhl, Mattacola and Malone (2005) reported low to fair inter-rater agreement for two novice investigators rating movement quality during a single leg squat in uninjured young adult participants. Ratings were based on criteria including the hip and knee position with explicit reference angles used to rate each segment. Ratings for each segment were then combined to give an overall score. The results of this study suggested that providing reference angles for visual rating of movement quality did not aid inter-rater agreement and this has been supported by other authors (Knudson, 1999). The percentage agreement reached (67 and 71%) was certainly no better than in our study where no reference angles were provided. Better inter-rater agreement (κ = 0.77 to 0.80) was reported in the study by Ekegren et al. (2009). A further study by Piva et al. (2006) also reported moderate levels of inter-rater agreement (percent agreement: 80%) when visually rating movement quality, during a lateral step down task, in participants with patellofemoral dysfunction. This level of agreement is similar to the highest levels of inter-rater agreement reached in the current study, however we emphasise the need for caution in this comparison due to the injured nature of the population in this study. Most ratings in the Piva et al. (2006) study were dichotomous, based on five criteria including the position of the trunk, pelvis and knee, and ratings were also combined to arrive at a final score which quantified the movement quality during the tests as poor, medium or good. We note that in these previous studies reporting inter-rater agreement the maximum number of raters was three. Our current study appears to be one of the few, if any studies, reporting inter-rater agreement in this area with a higher number of raters.

The Experienced physiotherapists mostly reached higher levels of intra-rater agreement than the Novice and Inexperienced physiotherapists and it is likely this difference in agreement was clinically

important (greater than the smallest worthwhile difference, Cohen 0.2). However, the lesser experience of the Novice physiotherapists did not improve intra-rater agreement compared to the Inexperienced physiotherapists. Although less likely to be clinically important, there is some evidence inter-rater agreement improves with experience. Brunnekreef et al. (2005) reported that experienced clinicians demonstrated higher levels of agreement and showed less variation between ratings when visually rating gait in participants with orthopaedic impairments. Furthermore von Porat, Holmstrom and Roos (2008) concluded that the experienced nature of the physiotherapists in their study improved agreement when rating knee stiffness during five functional tests while there were low levels of agreement for novice athletic trainers rating single leg squat in the study by DiMattia et al. (2005). In contrast the level of experience was not a factor in the visual rating of performance during a vertical jump test by kinesiology students and professors in the study by Knudson (1999). Knudson (1999) suggested that rating ability may be more related to the perceptual style of the raters or specific training in the rating task.

For all physiotherapists, intra and inter-rater agreement based on Dichotomous ratings was better than agreement based on Ordinal ratings. Consistent with this difference based on the AC1 statistic, the percentage agreements also appeared better with dichotomous ratings. This was in keeping with our expectations and this difference in agreement is very likely to be clinically important. Other authors have suggested that rating scales with fewer categories should produce higher agreement (Chmielewski, et al., 2007; Ekegren, et al., 2009). Thus dichotomous scales may be most appropriate for visual ratings that do not need more sensitive scales to quantify movement dysfunction (Chmielewski, et al., 2007). In contrast the method of rating movement quality (Segment versus Overall) did not appear to consistently influence the level of agreement achieved. This was a surprise as we expected the Segment method (with some guidance provided as to how to rate movement quality) to produce higher levels of agreement. This is however in accordance with conclusions reached by Chmielewski et al. (2007) who also failed to show a consistent benefit of a segment approach. It may be that visual rating of movement quality should be done using an overall approach and clinically this is probably not uncommon. It is likely that clinically this is the most time efficient approach while still providing adequate information. Knudson (1999) provided some support for this

view suggesting that a rating based on an overall impression of whole body motion during a task like a vertical jump was more reliable than ratings of individual segments.

In designing our study we decided not to give detailed instructions or training to the physiotherapists on how to make ratings. This was purposefully done as we wanted to investigate the level of agreement that was likely to exist in current clinical practice. It's possible that more explicit instructions or examples of what constituted each level of movement dysfunction would have increased agreement. It is also interesting to note that the knee segment, with the most explicit criteria for rating movement quality (anatomical reference of the patella relative to the second toe), did appear to produce higher agreement than the other body segments. Chmielewski et al. (2007) concluded that rating scales should include stricter criteria with more explicit instructions to improve agreement between raters.

As suggested by Chmielewski et al. (2007) the main clinical concern in evaluating agreement during visual rating is to avoid ratings that would result in different clinical decisions. As visual observation of lower extremity functional tests is one of many assessments (alongside for example strength and flexibility) used in making a clinical decision, we suggest the agreement reached (particularly by Experienced physiotherapists based on Dichotomous ratings) was acceptable for clinical use. Better agreement (particularly inter-rater) would be preferable if the functional tests were to be used in a mass screening situation where other assessments were not being carried out. Experienced physiotherapists should be used in such situations to help avoid inappropriate clinical decisions. However improved levels of agreement are probably desirable and this should be achievable with further development of rating scales using as few categories as considered necessary to clinically quantify movement quality. Further studies are also required comparing kinematic analysis to visual ratings to assess the validity of visual ratings of movement quality. Only a few studies appear to have looked at this to date. Krosshaug et al. (2007) concluded that visual estimation of joint angles during running and cutting must be interpreted with caution. More recently Ekegren et al. (2009) found that observational risk screening used to rate risk of anterior cruciate ligament injury, based on a drop-jump landing, showed acceptable specificity but not sensitivity.

The major limitations of this study are the relative homogeneity and small number of healthy individuals who were rated, that ratings were made from an anterior view only and the different rating environment of the experienced physiotherapists. Ratings of a clinical population may have given us an indication of agreement across a wider range of movement patterns and this is an area that needs further investigation. We emphasise that the agreement reported in this study only pertains to healthy populations and may not represent agreement when rating populations with lower extremity dysfunction. Further studies are required to assess the reliability of these tests in identifying readiness for return to sport and progress during rehabilitation with injured populations. However, visual rating of movement quality is used clinically as part of a screening process for risk of injury and our results are likely to give an indication of agreement in these circumstances. Additionally we caution that our method of calculating agreement did not preserve complete independence of ratings and while we consider it unlikely this may have inflated the AC1 values reported. Further studies are required where larger numbers of individuals are rated by similar numbers of physiotherapists.

As suggested by Hopkins et al. (2009) when considering the sample size we prefer to look at adequate precision based on the confidence interval rather than adequacy of power for statistical significance. Based on this approach a bigger sample size would have been ideal however we were constrained by resource issues. Precision for an agreement coefficient is thought to be adequate when the uncertainty in the estimate (represented by its confidence interval) does not span more than two qualitative magnitude thresholds (Hopkins & Manly, 1989). Based on the scale proposed by Landis and Loch (1977) the majority of the confidence intervals for our estimates of inter-rater agreement meet this requirement and thus suggest adequate precision. However we acknowledge there is less certainty in the reported intra-rater agreement (as evident by the range) but still feel the estimates provide a useful guide to agreement, indicating at least fair agreement for most ratings.

In this study only video from an anterior view was utilised as this appears common in the literature and also clinically for rating frontal plane control. It is possible that this 2D projection of a movement pattern may not be an adequate representation of movements in some planes or an ideal simulation of what is done clinically. Furthermore viewing of the transverse plane motions included on our rating scale may have been difficult when presented with only an anterior view. While we believe this is not

uncommon clinically it is possible that other views could be used and this may have an effect on the level of agreement. Additionally for the trunk and pelvis we cannot be sure if the agreement we have reported is based on frontal or transverse plane observations as our scale combined the two. Furthermore to mimic clinical practice we allowed physiotherapists to view as many repetitions of the video clips as they wished and this resulted in a lack of standardisation in this regard. As a result all agreements reported are not based on the same number of observations. We also acknowledge that the presence of anatomical markers may have improved agreement. As the markers were small we feel this effect is likely to have been negligible, however, further research without markers is needed to quantify the effect of markers on reliability of ratings.

Finally although we had a different rating environment for the experienced physiotherapists, viewing time was very similar and a recent study suggests ratings made in a group situation from a large screen are comparable to those made by the same individuals on a small screen (Ekegren, et al., 2009). We did however rely on the experienced physiotherapists self report of time spent making ratings and this may not have been completely accurate. Additionally as we only used physiotherapists in the current study the agreement reported cannot be generalised to other practitioners who may perform visual ratings.

Conclusion

Physiotherapists' visual rating of movement quality via video during four lower extremity functional tests resulted in slight to almost perfect intra-rater but only fair to good inter-rater agreement. The best agreement was achieved by the dichotomously classified ratings of experienced physiotherapists, irrespective of whether a segment or overall rating approach was used. Increased experience was likely to result in a clinically important improvement in intra-rater agreement but not inter-rater agreement.

Supplementary information

Statistical analyses

An Excel spreadsheet (Hopkins, 2007) was used to compare the intra-rater agreement of the three physiotherapy groups (Inexperienced, Novice and Experienced). To assess these differences using the Segment rating method the mean of all segments' AC1 was used. Fisher transformation of the AC1 was used for all comparisons of any two group means (e.g. Experienced versus Novice). The spreadsheet is based on the assumption that the Fisher z transformation of the AC1 like that of all correlation coefficients has a normal distribution. This allows the uncertainties in each estimate of agreement (confidence limits) to be combined in the same manner as independent variances followed by back transformation. The spreadsheet provided the likelihood of the mean difference in AC1 being greater than the smallest worthwhile difference (in standardized Cohen units equal to 0.2 of the unweighted mean of the between-subjects standard deviations of the two groups) and thus clinically important. Differences in inter-rater agreement between groups were similarly estimated using an additional Excel spreadsheet (Hopkins, 2006). For inter-rater agreement the smallest worthwhile difference in AC1 (threshold for clinical importance) was set at Fisher 0.1. These two spreadsheets were also used in the same manner to assess the effect of the Dichotomous ratings versus Ordinal ratings on level of intra and inter-rater agreement.

$$AC1 = \frac{p_a - p_{ey}}{1 - p_{ey}}$$

$$\pi_q = \frac{1}{n} \sum_{i=1}^{n} \frac{r_{iq}}{r}$$

$$p_{ey} = \frac{1}{Q-1} \sum_{q=1}^{Q} \pi_q (1 - \pi_q)$$

$$p_a = \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{q=1}^{Q} \frac{r_{iq}(r_{iq}-1)}{r(r-1)} \right\}$$

Figure 4.2:   Formula for calculating the AC1 in the general situation of more than two raters and a rating scale with greater than two categories (Blood & Spratt, 2007)

r_{iq} indicates the number of raters who classified the ith object into the *qth category. The index* i ranges from 1 to *n* and q ranges from 1 to Q, where n is the number of objects rated and Q is the number of categories in the rating scale. r indicates the total number of raters. For use in the p_{ey} formula, π_q needs to be calculate, the probability that a rater classifies an object into a category *q. The* y in the p_{ey} formula indicates that this calculation takes into account the probability of a random rating (*y*). These formulas allow the AC1 calculation (Blood & Spratt, 2007).

CHAPTER 5: KINEMATICS DURING LOWER EXTREMITY FUNCTIONAL SCREENING TESTS IN YOUNG ATHLETES – ARE THEY RELIABLE AND VALID?

Overview

Purpose: To determine if lower extremity functional tests are reliable and valid screening tests of lower extremity dynamic alignment in healthy young athletes.

Methods: Peak three-dimensional pelvis and lower extremity kinematics were quantified in 23 uninjured young athletes (11 ±1 years) during three lower extremity functional tests (Small Knee Bend [SKB], Single Leg SKB and Drop Jump) and Running. A nine camera motion analysis system captured three trials of all tests. All functional tests were repeated by 10 young athletes eight to ten weeks later. Visual 3D and Labview were used to process all data. Intraclass correlation coefficients (ICC) and typical errors (TE) were used to assess within- and between-day reliability of all variables. Pearson correlation coefficients were used to evaluate associations between peak kinematics during the Small Knee Bend and Drop Jump and between the Single Leg SKB and Running.

Results: Within-day reliability was excellent (ICC ≥ 0.85) and between-day reliability was excellent to good (ICC range 0.60 to 0.92) for the majority of kinematic variables. Correlations for peak lower extremity kinematics between Small Knee Bend and Drop Jump were moderate to very large (r = 0.39 to 0.87) as were correlations between Single Leg SKB and Running (r = 0.45 to 0.84).

Conclusions: Small knee bend lower extremity functional tests are a useful clinical tool for assessing dynamic lower extremity alignment in healthy young athletes.

Background

Injuries are a significant problem for child and adolescent athletes (Adirim & Cheng, 2003). Injury prevention in youth sport is becoming a public health priority Emery (2010) and pre-participation screening of young athletes has been recommended to help reduce injury risk (Caine & Golightly, 2011). In youth sports the ankle and knee are frequently injured (Caine, Caine, & Maffulli, 2006), with overuse injuries increasingly prevalent due to increased exposure and specialisation (Caine, Maffulli, & Caine, 2008). Dynamic alignment/control of the lower extremity under load is considered a key risk

factor for lower extremity injuries (Powers, 2010; Reiman, et al., 2009) and has been recommended as a component of lower extremity screening (Caine & Golightly, 2011; Mottram & Comerford, 2008; Reiman, et al., 2009). The reliability of screening tests is important but has not been reported for young athletes completing many lower extremity screening tests. The drop jump is the most common test reported in the literature for screening lower extremity alignment/control in young athletes (Ford, et al., 2007; Noyes, et al., 2005). Other different but similar lower extremity screening tests are known by various names including small knee bend, partial squat, mini squat, single limb mini squat and single leg squat. To avoid confusion this paper refers to two small knee bend lower extremity functional tests; one on double leg support referred to as the Small Knee Bend (SKB) and the other performed on single leg support referred to as the Single Leg SKB. The SKB and Single Leg SKB may be useful for screening young athletes but have received little attention. Örtqvist et al., (2011) concluded that a single limb mini squat test should be included in clinical evaluation of dynamic knee position in children. Further development of easily administered, reliable screening tests for young athletes will assist in the enhancement of injury prevention and rehabilitation programmes.

Previous studies have shown that two dimensional (2D) measures of valgus alignment and frontal plane knee excursion in young athletes during a drop jump are reliable (ICC≥0.90) (Noyes, et al., 2005; Sigward, Ota, & Powers, 2008). In a study of young soccer players, the majority of three dimensional (3D) kinematic variables were also reliable (within-day ICC≥0.93, between-day ICC≥0.60) (Ford, et al., 2007). Reliability of the pelvis during lower extremity functional tests has however not been reported and Ford et al., (2007) has produced the only study to have assessed reliability over longer than one week (seven weeks). Although reliability over one week may be appropriate for screening purposes, extended periods are needed to ensure the stated reliability has relevance to the period of sporting seasons, rehabilitation timeframes and longitudinal studies. The length of a sporting season for a young athlete will vary depending on the sport, however 2-3 months is not uncommon. In combined biomechanical and epidemiological studies, biomechanical variables measured at the start of the season need to demonstrate sufficient reliability over the length of the season if they are to be considered valid risk factors for subsequent injuries recorded during the season (Ford, Myer, & Hewett, 2007). Given the developmental changes occurring in young athletes it may be that substantial variation in movement is to be expected over periods of 2-3 months and thus such

measures are not sufficiently reliable for combined biomechanical and epidemiological studies. Additionally Hopkins (2000) suggests the typical error used for estimates of sample size and individual differences in experiments needs to come from a reliability study of the same duration as the experiment.

The Small Knee Bend (Sahrmann, 2002), single leg squat (Stensrud, et al., 2011) and single leg mini-squat (Ageberg, et al., 2010) have been used in older populations to assess dynamic lower extremity alignment/control. The Small Knee Bend (also known as a partial squat or mini-squat) is described by Sahrmann (2002) as part of a lower quarter examination. Single leg variations of the test have also been described (Mottram & Comerford, 2008; Reid, et al., 2003). The tests are commonly used by physiotherapists and sports physicians to assess dynamic trunk, pelvic and lower extremity alignment/control with results aiding clinical decisions when considering risk of injury, prescription of exercises and also evaluating progress during rehabilitation. As the small knee bend tests are performed more slowly than the drop jump they may be easier to assess clinically via observation. The Single Leg SKB also offers the opportunity to assess one leg at a time which is likely to be relevant to performance in young athletes and allows comparison of sides for clinical assessment when an athlete presents with unilateral symptoms.

A key component of small knee bend screening assessment is alignment of the knee relative to the foot in the frontal plane. Clinical observation of the patella position relative to a line extending vertically from the 1st or 2nd toes (or more generally the knee medial to the foot) is a common measure of frontal plane knee control (Ageberg, et al., 2010; Bell, et al., 2008; Padua, et al., 2009; Trulsson, Garwicz, & Ageberg, 2010). A knee position medial to the foot during functional activities involving hip and knee flexion is more common in individuals with patellofemoral dysfunction (Willson & Davis, 2008a) or ACL injury (Hewett, et al., 2005). Poor dynamic alignment has also been described as a combination of excessive pelvic drop, hip adduction and internal rotation, knee abduction, tibial internal or external rotation and foot hyperpronation (Earl, et al., 2005; Powers, 2003; Sahrmann, 2002). Poor dynamic alignment/control during activities such as running, squatting and landing, is considered a key risk factor for the development of common lower extremity overuse injuries (Boling, et al., 2009; Powers, 2010).

Our current study builds on our previous work (Whatman, et al., 2011a) that reported peak kinematics during the loading phase of various lower extremity functional tests (including the SKB and Single Leg SKB) are reliable in adults both within- (ICC ≥ 0.92) and between-days (ICC ≥ 0.80). We also reported large to very large associations between peak kinematics during SKB and Single Leg SKB and peak kinematics during running in an adult population. Clinicians often assume functional tests are an effective method of diagnosing actual movement dysfunction. However evidence for the validity of using slower, lower load activities such as the SKB or Single Leg SKB to predict movement during faster, higher load activities is limited and requires further investigation across various ages.

Small knee bend lower extremity functional tests have the potential to be used as a screening tool in young athletes, however little is known about the typical variation in 3D kinematic variables in this age group. If tests are to be used with young athletes, physiotherapists need to know if the kinematics are consistent enough from day to day for making clinical decisions. The association between the kinematics recorded during small knee bend functional tests and those occurring during actual function (e.g. running and drop jump landing) in young athletes, also warrants investigation.

Purpose

To determine if small knee bend tests are reliable and valid screening tests of lower extremity dynamic alignment in healthy young athletes. Specifically to investigate (1) the within- and between-day reliability of peak 3D pelvis and lower extremity kinematics during three lower extremity functional tests (SKB, Single Leg SKB and Drop Jump) and (2) the associations between peak kinematics for dual limb tasks (SKB and Drop Jump) and single limb tasks (Single Leg SKB and Running).

Methods

Participants

Twenty three participants (mean ±SD: 11 ±1 y, 153 ±10 cm, 44 ±8 kg) with no musculoskeletal problems volunteered for this study. All participants were part of a structured long term athlete development programme and competed in a variety of sports. The study was approved by the

university ethics committee. All participants/parents received verbal and written information regarding the study and all participants gave assent prior to participation. All parents also provided written informed consent prior to testing.

Instrumentation

A nine camera motion analysis system (Qualysis Medical AB, Sweden) sampling at 240 Hz collected kinematic data. Participants' pelvis and both legs were instrumented with retro-reflective markers (19 mm diameter) secured to anatomical locations (sacrum, bilateral ASIS's, iliac crests, greater trochanters, medial and lateral femoral epicondyles, mid-patella, medial and lateral malleoli, head of 5th metatarsal, head of 1st metatarsal, posterior calcaneus) by an experienced musculoskeletal physiotherapist (Figure 5.1). The foot markers were placed over shoes as the participants were part of a bigger study investigating running gait which required shoes to be worn. Standardisation of shoes would have been ideal however this was not possible and as we were only interested in within participant measures (not absolute comparisons between participants) we didn't consider it essential. Four cluster marker sets were attached to the thigh and shank of each leg (Figure 5.1). This is considered more accurate and practical for tracking motion than individual skin markers (Angeloni, et al., 1993). The anatomical markers were used for construction of a skeletal model using Visual 3D (C-Motion Inc, USA).
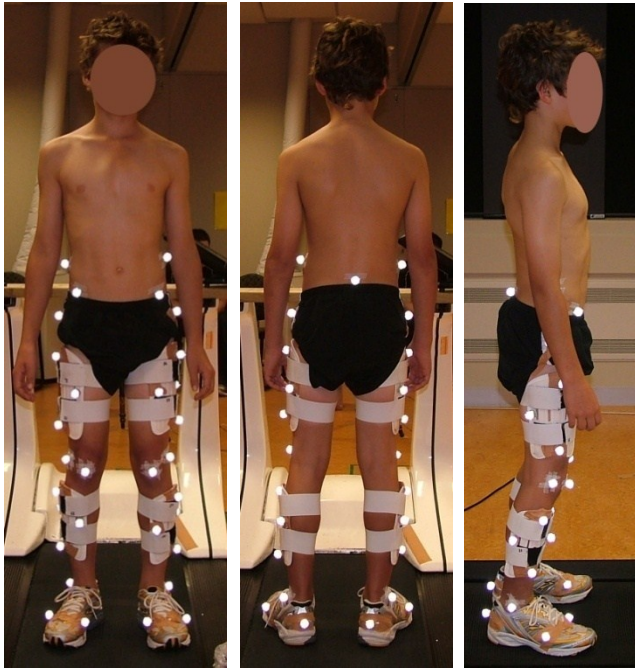
Figure 5.1: Participant instrumented with markers (note medial knee and ankle anatomical markers used for skeletal model construction are not shown as they were removed during functional tests).

Testing protocol

All twenty three participants attended the Motion Analysis Laboratory on one occasion. Ten participants (43%) returned for repeat testing eight to ten weeks later. This is similar to the sample size used in previous studies using similar methods, variables and time interval between tests (Ford, et al., 2007). Following instrumentation of the retro-reflective markers a static standing trial was collected.

The order of the three functional tests was randomized among participants. The Single Leg SKB was performed for each leg by all participants. For all tests participants were given standardized verbal instructions prior to each test (Table 5.1) and the researcher demonstrated each test. Participants were required to keep their heels on the ground throughout each of the small knee bend tests to assist in standardizing the range of hip and knee flexion without the need for additional monitoring and equipment. This procedure along with the simple instructions maximizes the clinical utility of the tests. Prior to the tests participants walked a few steps to move into the test area of the laboratory and

take up their natural stance position. Practice for all tests was allowed until the researcher was confident the test was performed consistently (this usually required 3-5 practice attempts). The velocity of each test was set by a three second count made by the researcher (verbally counted as "one and two and three") during the knee flexion phase of each test. During all tests participants were instructed to maintain visual focus on a cross positioned on a wall directly in front of them at eye level to maintain an upright trunk position (Whatman, et al., 2011a) improving consistency and simulate a functional head and trunk position. For the Drop Jump participants started on a 25 cm high box and dropped directly down off the box onto the force plate and immediately jumped vertically as high as possible. Participants performed three repetitions of all tests and were also recorded running on a treadmill for 30 s at a self selected velocity (mean ±SD = 2.2 ±0.2 m/s). All participants were familiar with running on a treadmill as part of their training programme prior to this study. Treadmill running kinematics in young athletes have previously been reported as reliable (e.g. ICC = 0.51 to 0.92. for knee abduction) (Sheerin, Whatman, Hume, & Croft, 2010).

Table 5.1:   Description of the small knee bend functional tests used in the study.

Functional test	Test description
Small Knee Bend (SKB)	Starting from a standing position, participants performed a partial squat (hip and knee flexion) with the trunk maintained in an upright position while looking straight ahead. Participants were instructed to continue the SKB until they reached maximum dorsiflexion without lifting their heels and then return to upright standing.
Single Leg SKB	Standing on one leg, with the contralateral hip in neutral and knee flexed to approximately 80°, participants performed a SKB as described above.

Data processing

All static and motion (functional test and running) trials were tracked using the Qualysis motion capture software and exported to Visual 3D (C-Motion Inc, USA). In Visual 3D the rigid link model (pelvis, thigh, shank and foot) created from the static file was assigned to all imported motion files to calculate kinematic data. Data were filtered with a second-order Butterworth bidirectional low-pass

filter with a cut-off frequency of 6 Hz. The Cardan sequence y-x-z was used for the calculation of joint angles which was equivalent to flexion/extension, abduction/adduction, axial rotation and in this case equivalent to the Joint Coordinate System described by (Grood & Suntay, 1983). The only exception to this was for calculation of pelvic angle with respect to the laboratory where we used the sequence z-x-y (axial rotation, obliquity, tilt) which is recommended in the Visual 3D documentation based on the suggestions of Baker (2001).

All pelvis and lower limb kinematic data were exported to a customised Labview programme and processed to provide peak joint angles during the loading phase (knee flexion phase) of each functional test and the right and left stance phase (start to maximum knee flexion only) of running. The loading phase in the Drop Jump was from initial ground contact (determined by the force plate recordings based on the onset of the vertical ground-reaction force) to maximum knee flexion and for the small knee bend tests from minimum to maximum knee flexion. The mean of each variable, across the three repetitions of each functional test, was used in between-day statistical analyses and for comparisons between the SKB and Drop Jump. The mean of ten running strides was used for the comparison between running and Single Leg SKB. The maximum medial position of the patella marker in the frontal plane (y coordinate in the laboratory coordinate system) relative to the 1st MTP marker, termed medial knee displacement (MKD), was analysed in the same manner as for the joint angles.

Statistical analyses

Proc mixed in Statistical Analysis Systems (Version 9.1, SAS Institute, Cary, NC) was used to estimate within-subject trial to trial variability for within-day and between-days. Trial to trial variability within-day was the residual error in the model. The mixed model estimated residuals for day 1 and day 2, a random effect for Subject by Day, and fixed effects (interaction of Day and Trial) to allow for a change in the mean between all trials on both days. Within- and between-day reliability was expressed as a typical error (TE) in degrees and intra-class correlation coefficients (ICC). Typical error was interpreted as the expected variation in peak joint angle when one participant was tested on

repeat occasions. The ICC classifications of Fleiss (1999) were used to describe the magnitude of ICC values (<0.4 = poor, 0.4 to 0.75 = fair to good, >0.75 = excellent).

Using data from initial testing only (all 23 participants), Pearson correlation coefficients were calculated to assess, for peak joint angles, the magnitudes of the associations between the single limb tasks (Single Leg SKB and running) and the dual limb tasks (SKB and Drop Jump). The magnitudes of these correlations were described as trivial (0.0-0.1), small (0.1-0.3), moderate (0.3-0.5), large (0.5-0.7), very large (0.7-0.9), or extremely large (0.9-1.0) (Hopkins, et al., 2009).

Results

For the majority of kinematic variables the SKB and Single Leg SKB were reliable tests of lower extremity dynamic alignment in the young athletes. Based on three trials, the within-day ICC values for all variables (ICC = 0.67 to 0.99; Table 5.2) were generally higher than the between-day ICC values (ICC = 0.26 to 0.92; Table 5.3). Within-day ICC's were all greater than 0.80 (majority greater than 0.9), except for pelvic rotation and left ankle dorsiflexion during the Drop Jump (ICC = 0.67 to 0.73). Within-day typical errors for all variables, representing the typical variation for an individual participant on repeat tests, ranged from 0.4° to 3.6°. For most variables the within-day ICC and typical error range across all tests was small indicating similar reliability irrespective of the test.

Table 5.2:   Within-day reliability (Typical Error; Intraclass Correlation Coefficient) of the peak angle (°) or medial knee displacement (cm) for all three functional tests for 23 young athletes.

		Drop Jump		Small Knee Bend		Single Leg SKB	
		Right	**Left**	**Right**	**Left**	**Right**	**Left**
Pelvis	Lateral tilt°	1.1, 0.88	1.0, 0.88	0.6, 0.94	0.9, 0.86	1.5, 0.94	1.0, 0.87
	Rotation°	1.7, 0.72	2.1, 0.67	1.1, 0.92	1.1, 0.92	1.6, 0.82	1.4, 0.88
Hip	Flexion°	2.8, 0.95	3.3, 0.92	3.2, 0.97	3.0, 0.97	2.6, 0.97	2.2, 0.97
	Adduction°	1.6, 0.86	1.4, 0.91	0.7, 0.95	0.8, 0.94	1.4, 0.95	1.4, 0.94
	Internal rotation°	1.4, 0.93	1.2, 0.97	1.1, 0.93	0.9, 0.98	0.9, 0.98	1.1, 0.98
Knee	Flexion°	2.9, 0.95	2.6, 0.94	1.9, 0.99	1.9, 0.84	1.6, 0.98	1.6, 0.97
	Abduction°	1.3, 0.93	1.4, 0.96	0.4, 0.99	0.7, 0.97	0.6, 0.98	0.7, 0.97
Ankle	Dorsiflexion°	3.6, 0.99	2.7, 0.70	1.1, 0.99	0.7, 0.99	0.9, 0.99	0.9, 0.96
	Eversion°	3.2, 0.73	1.6, 0.86	0.9, 0.98	0.6, 0.98	0.7, 0.99	0.5, 0.99
Knee	Medial knee displacement (cm)	0.9, 0.81	0.8, 0.90	0.4, 0.96	0.5, 0.95	0.6, 0.94	0.5, 0.90

Typical Error 90% CL ~×/÷1.47, lowest ICC 90% CL ~ ±0.35, highest ICC 90% CL ~ ±0.02. SKB=small knee bend.

The majority of between-day ICC's for Single Leg SKB were greater than 0.74 and typical errors ranged from 1.2° to 5.7°. For the SKB most between-day ICC's were greater than 0.60 and typical errors ranged from 0.9° to 10.4°. The Drop Jump generally had worse between-day reliability with several ICC's lower than 0.60. The worst between-day ICC's of 0.26 to 0.27 were still associated with reasonably small typical errors (3 to 4° or 1 to 2 cm).

Table 5.3:   Between-day reliability (Typical Error; Intraclass Correlation Coefficient) of the peak joint angle (°) and medial knee displacement (cm) for all three functional tests for ten young athletes.

		Drop Jump		Small Knee Bend		Single Leg SKB	
		Right	**Left**	**Right**	**Left**	**Right**	**Left**
Pelvis	Lateral tilt°	2.3, 0.45	1.8, 0.67	1.6, 0.61	1.1, 0.87	2.2, 0.80	2.3, 0.67
	Rotation°	2.4, 0.40	2.7, 0.44	2.6, 0.77	2.3, 0.64	2.4, 0.66	1.3, 0.81
Hip	Flexion°	7.3, 0.63	7.7, 0.56	10.4, 0.79	9.6, 0.83	4.8, 0.88	5.7, 0.82
	Adduction°	1.4, 0.89	3.2, 0.56	1.2, 0.69	2.5, 0.67	2.8, 0.86	2.4, 0.87
	Internal rotation°	4.6, 0.27	5.5, 0.43	2.3, 0.68	3.0, 0.69	1.8, 0.91	2.6, 0.87
Knee	Flexion°	5.4, 0.79	4.8, 0.79	7.6, 0.87	6.5, 0.91	5.3, 0.74	4.1, 0.92
	Abduction°	3.4, 0.54	4.3, 0.65	1.4, 0.88	1.8, 0.86	2.0, 0.74	1.4, 0.90
Ankle	Dorsiflexion°	4.6, 0.51	3.2, 0.55	2.5, 0.75	3.4, 0.71	2.1, 0.90	2.4, 0.78
	Eversion°	3.2, 0.64	3.3, 0.26	2.5, 0.41	3.6, 0.40	2.5, 0.77	2.0, 0.54
Knee	Medial knee displacement (cm)	1.5, 0.52	2.1, 0.26	1.4, 0.60	0.9, 0.84	2.1, 0.27	1.2, 0.51

Typical Error 90% CL~×/÷1.47, lowest ICC 90% CL~±0.53, highest ICC 90% CL~±0.11. SKB=small knee bend.

SKB was a valid screening test of lower extremity dynamic alignment in young athletes given there were large to very large correlations between peak hip, knee and ankle angles during the SKB and the Drop Jump (r = 0.57 to 0.87; Table 5.4). Hip internal rotation showed the strongest correlations (r = 0.82 and 0.87). Medial knee displacement during the SKB was moderately correlated to medial knee displacement during the Drop Jump. Excluding medial knee displacement, confidence limits indicated the true correlations were very likely to be at least moderate (≥ 0.3) and possibly large (≥ 0.5). There were also moderate to very large correlations between peak hip, knee, ankle angles and medial knee displacement during the Single Leg SKB and those recorded during the loading phase of running (r = 0.45 to 0.84; Table 5.5). Based on the confidence limits, all correlations were likely to be at least moderate and several were likely to be large including knee abduction and ankle eversion.
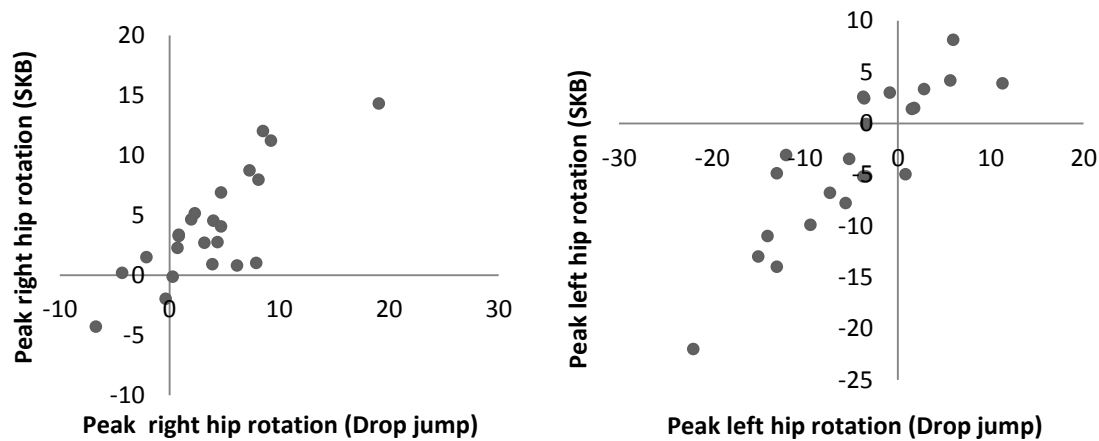
Figure 5.2: Example scatter plots showing the association between Small Knee Bend (SKB) and Drop Jump for peak hip rotation angles (°) (R leg -ve = external rotation, L leg -ve = internal rotation). Each point represents the mean of three trials for a given participant.

Table 5.4:   Associations between peak joint angles during the Small Knee Bend and Drop Jump and between the Single Leg SKB and Running expressed as Pearson correlation coefficients (90% CL).

		Small Knee Bend versus Drop Jump		Single Leg SKB versus Running	
		Right leg	**Left leg**	**Right leg**	**Left leg**
Hip	Adduction°	0.61 (0.33-0.79)	0.57 (0.27-0.77)	0.45 (0.10-0.70)	0.57 (0.24-0.79)
	Internal rotation°	0.82 (0.65-0.91)	0.87 (0.74-0.93)	0.46 (0.12-0.70)	0.67 (0.39-0.84)
Ankle	Eversion°	0.71 (0.46-0.85)	0.66 (0.39-0.82)	0.78 (0.56-0.90)	0.65 (0.35-0.83)
Knee	Abduction	0.60 (0.32-0.79)	0.63 (0.36-0.81)	0.64 (0.36-0.81)	0.84 (0.68-0.93)
	Medial knee displacement (cm)°	0.42 (0.08-0.67)	0.39 (0.03-0.67)	0.46 (0.11-0.70)	0.53 (0.19-0.75)

Discussion

The SKB was a reliable screening test of lower extremity dynamic alignment in young athletes. The within-day reliability of the majority of kinematics was excellent and the between-day reliability was good to excellent. To our knowledge this is the first study to report 3D kinematic reliability of the pelvis, hip, knee and ankle during these specific functional tests in this age group. Previous studies of young athletes have reported the reliability of hip, knee and ankle kinematics in the drop jump only

(Ford, et al., 2007). We believe that all these tests could be useful in screening lower extremity dynamic alignment. Compared to the drop jump, the small knee bend tests allow screening of single leg function and may be easier to assess via clinical observation.

Within-day reliability was better than between-day for all tests which supports the findings of previous studies (Ford, et al., 2007; Kadaba, et al., 1989), including our previous study investigating the reliability of small knee bend tests in adults (Whatman, et al., 2011a) where the majority of kinematics of interest showed acceptable reliability (within-day ICC ≥ 0.92; between-day ICC ≥ 0.80). The reliability in our current study on young athletes was similar for the majority of the same peak joint angles. However the typical errors (TE) for hip and knee flexion in our young athletes were higher than those of the adults in our earlier study (hip = 2.8º to 5.4º; knee = 2.4º to 3.7º). These results suggest young athletes' lower extremity movements are not as reliable in the sagittal plane compared with adults. Alternatively this additional variability could represent errors in marker placement between days (Kadaba, et al., 1989) and/or increased variability over longer timeframes. Due to the long duration between repeat tests, marking the skin to improve accuracy of marker placement was not possible in this study. Despite standardised marker placement by the single investigator there was some variability in our static trials (typical error ~ 1 to 2º) which is likely a combination of marker placement error and variations is static posture in the young athletes. Variations in static alignment and neuromuscular development in young children may contribute to variability in repeat motion analysis testing (Ford, et al., 2007). However peak angles in the frontal and transverse planes showed similar reliability to our previous study and these are of most importance when screening lower extremity alignment/control as they are the movements most commonly linked to injury. The tests should also be useful in longitudinal studies investigating the effects of interventions designed to improve lower extremity alignment/control in young athletes. Consideration could be given to improving consistency of movement in the sagittal plane as this has been suggested to influence dynamic control in the other planes (Powers, 2010).

Reliability of the small knee bend tests in the young athletes was better for most peak joint angles than the Drop Jump. The reliability of peak hip, knee and ankle 3D joint kinematics in young soccer players when landing from a drop jump has been reported to be reliable for the majority of kinematics (within-day ICC = 0.93 to 0.99, between-day ICC = 0.60 to 0.92) (Ford, et al., 2007). This is similar to

the reliability of the Drop Jump in the current study for the majority of hip, knee and angle peak angles. Ford et al., (2007) did not report right and left legs separately but combined sides to report a single reliability value. Excluding hip flexion, the typical errors reported by Ford et al., (2007) are similar to those in our young athlete study (within-day 0.9º to 3.2º, between-day 1.3º to 5.5º). Other studies of older populations have reported comparable reliability during lower extremity functional tests such as the drop jump and step-down (Earl, et al., 2007; Kernozek, et al., 2005).

From a clinical view point the good to excellent between-day reliability of the peak hip joint kinematics in the small knee bend tests is important. The role of hip motion in the frontal and transverse planes has been a focus of many recent studies investigating lower extremity alignment and risk of injury (Cowan, et al., 2008; Reiman, et al., 2009). Powers (2010) suggested that altered hip kinematics may adversely affect tibiofemoral and patellofemoral joint function. Altered hip function may lead to a knee position medial to the foot during functional activities which is more common in individuals with patellofemoral dysfunction (Willson & Davis, 2008b). Of further clinical relevance is the reliability of the pelvis which to our knowledge has not been reported previously in young athletes. Within-day and between-day reliability of the pelvis was mostly similar to the other joints. Powers (2010) argued that intervention programmes to address proximal dysfunction related to the risk of knee injury should address both hip and pelvis control.

In addition to joint angles, we investigated the reliability of medial knee displacement as this is commonly used by physiotherapists as a means of visually assessing dynamic lower extremity alignment. The excellent within-day reliability (ICC = 0.81 to 0.96) was better than between-day reliability (ICC = 0.26 to 0.84). The poorer between-day reliability was likely affected by marking of the 1st metatarsal over shoes and not all participants remembering to bring the same shoes on day 2. We have some evidence to support marker placement error as there was only moderate between-day reliability of medial knee displacement, wearing shoes, in our static trials (ICC = 0.60). Maximum medial knee displacement has the potential to show higher reliability between-days and this is an area that requires further investigation. What remains unclear is whether or not the position of the patella is a good indicator of lower extremity 3D joint kinematics.

The typical errors reported in the current study give an indication to physiotherapists as to the absolute variation they can expect for any given athlete between consecutive tests. However visual observation is confounded by any variation observed visually being a composite (addition) of the typical errors reported in the individual planes. Interpretation of the ICC's alongside these typical errors provides evidence for physiotherapists that what they are observing on any given occasion is repeatable and representative of an athlete's kinematics. This is better than interpretation of the ICC's alone due to sensitivity to the heterogeneity of values between participants (Hopkins, 2000).

A limitation of our study is the uninjured population that we investigated. The reliability of these lower extremity functional tests in a population of young athletes with lower extremity injury, such as patellofemoral syndrome, needs further investigation. However the current results provide useful information for the use of the tests when attempting to predict future injury in a screening situation. Additionally precision for a correlation is thought to be adequate when the uncertainty in the estimate (represented by its confidence interval) does not span more than two qualitative magnitude thresholds (Hopkins & Manly, 1989). The lower between-day ICC's we reported did not have this level of certainty, so while they provide useful information they must be interpreted with caution.

The small knee bend tests were a valid screening test of lower extremity dynamic alignment in young athletes shown by the moderate to very large association between the peak 3D joint kinematics during the small knee bend tests and those occurring during running and landing for most variables of interest. Participants with higher peak angles in the functional tests had higher peak angles in running and landing. This association provides some preliminary support for the use of these tests as a screening tool in the clinic where visual observation of small knee bend tests may be easier than a drop jump or running. As the velocity of the movement during small knee bend test is slower it is likely that physiotherapists will be able to visually rate movement with greater reliability and validity. If an athlete has a current injury that prevents the use of a high load test such as a drop jump, an indication of alignment/control can still be gained via the use of the small knee bend tests. There are clinical advantages involving space and equipment which make the use of the Single Leg SKB test more feasible than direct assessment of running gait. We caution here that we are not suggesting the small knee bend tests can take the place of tests such as the drop jump or running assessment. We have

not performed a comparison of kinematics throughout (or at specific time points during) the small knee bend tests, drop jump and running, which would be required before this could be contemplated.

Conclusion

In healthy young athletes the reliability of peak 3D kinematics during the loading phase of small knee bend lower extremity functional tests is acceptable for the majority of kinematic variables of interest to physiotherapists. There was a moderate to large association between the Small Knee Bend and Single Leg SKB peak kinematics and those recorded during landing from a Drop Jump and running. Based on these results small knee bend lower extremity functional tests should be useful in helping physiotherapists make clinical decisions regarding lower extremity dynamic alignment and risk of injury in young athletes.

CHAPTER 6: THE RELIABILITY AND VALIDITY OF PHYSIOTHERAPIST VISUAL RATING OF DYNAMIC PELVIS AND KNEE ALIGNMENT IN YOUNG ATHLETES

Overview

Purpose: To investigate the ability of physiotherapists to visually rate dynamic pelvis and knee position in young athletes during lower extremity functional tests. *Methods:* Pelvis and knee alignment during lower extremity functional tests, in 23 athletes, was visually rated by 66 physiotherapists. Peak two-dimensional (2D) and three-dimensional (3D) lower extremity kinematics were also quantified. Physiotherapist ratings were compared to consensus visual ratings of an expert panel. The consensus ratings were also compared to peak kinematics. Physiotherapist experience and functional test velocity were assessed as factors affecting rating accuracy. Reliability of ratings was determined using percentage agreement (PA) and the first order agreement coefficient (AC1). Sensitivity, specificity and the diagnostic odds ratio (DOR) were calculated to assess the validity of ratings as well as differences in kinematics between groups based on the expert visual ratings. *Results:* Mean intra-rater agreement for all ratings was substantial (PA: 79% to 88%, AC1: 0.60 to 0.78). Inter-rater agreement ranged from fair to substantial (PA: 67% to 80%; AC1: 0.37 to 0.61). Sensitivity (≥80%) and specificity (≥50%) were acceptable for all tests except the Drop Jump. Experience (DOR 1.6 to 2.8 times better) and slower movement (4.9 times better) were possibly factors in better rating accuracy. Peak 3D and 2D kinematics were different between groups rated as having good versus poor alignment by the experts (likely to almost certainly), except for knee abduction angle. *Conclusions:* Visual rating by physiotherapists is a valid tool for identifying young athletes with poor frontal plane dynamic pelvis and knee alignment. Ratings are better with slower movements and possibly with increased clinical experience.

Background

The dynamic alignment of the lower extremity is considered a key risk factor for several injuries including patellofemoral pain (Powers, 2003) and iliotibial band syndrome (Reiman, et al., 2009). A number of authors have reported an association between dynamic alignment and current injury (Boling, et al., 2009; Souza & Powers, 2009; Willson & Davis, 2009). Common measures of dynamic

alignment include pelvic position and the medio-lateral knee position relative to the foot during functional tasks involving hip and knee flexion. These alignment measures are considered good indicators of movement quality (Sahrmann, 2002) with a knee position medial to the foot considered less than ideal. A neutral pelvic position in the frontal and transverse planes is also considered an important indicator of movement quality (Chmielewski, et al., 2007). Observation of pelvic position has been shown to contribute to the assessment of hip muscle function (Crossley, et al., 2011) and Reiman et al., (2009) concluded in a review that there was mounting evidence that hip weakness contributes to lower extremity injury across all ages. Given the importance of good pelvic and knee position the further development of clinically applicable techniques for identifying athletes with poor alignment is needed.

Techniques for assessing the dynamic position of the knee relative to the foot have been specifically noted as important in screening young athletes (Örtqvist, et al., 2011). With increasing rates of participation by children and adolescents in sport there is an increased risk of injury. The risk of overuse injury, particularly to the lower extremity is considered especially high in young athletes (Caine, et al., 2006). Reducing injury risk has therefore become a focus for the sports medicine community (Caine & Golightly, 2011). The key to reducing risk of injury and also preventing recurrence could be the identification of faulty or less than ideal movement patterns/alignment (Mottram & Comerford, 2008; Sahrmann, 2002). As a result many now suggest that young athletes should be screened for movement quality, so the development of screening tests is promoted (McLeod, et al., 2011).

The most common method used clinically to assess dynamic pelvis and knee alignment/control is visual observation of functional tests. Physiotherapists regularly use information gained from visual observation to aid in the clinical decision making process and to decide on appropriate intervention. Instantaneous feedback during training and rehabilitation, aimed at correcting faulty movement is also based on visual observation. The small knee bend (SKB) tests (Sahrmann, 2002) are common clinical lower extremity functional tests used in this screening/assessment process. These tests are easily administered in the clinic setting and used to replicate sporting/daily activities. Similar tests described by other authors with alternate names are the mini squat and single limb mini squat (Ageberg, et al.,

2010). In young athletes another functional test commonly reported in the literature, but not as commonly used in physiotherapy clinical practice, is the drop jump (Noyes, et al., 2005). Further validation of the use of functional tests in young athletes for the evaluation of movement quality by visual assessment is required.

Previous studies investigating the reliability of visual assessment of lower extremity functional tests have revealed inconsistent results (intra-rater = moderate to excellent; inter-rater = slight to good) (Ageberg, et al., 2010; Chmielewski, et al., 2007; Crossley, et al., 2011; DiMattia, et al., 2005; Ekegren, et al., 2009; Örtqvist, et al., 2011; Whatman, et al., 2011b). This is likely due to differences in the functional tests themselves as well as the rating methods, variations in the amount of rater training, raters with varied experience and differences in analyses. Few studies have reported the reliability of rating the pelvis and there is little known about the reliability of rating young athletes. The only study involving children reported moderate intra- and inter-rater reliability for ratings of medio-lateral knee position (Örtqvist, et al., 2011) and it has been identified that further studies involving children are required (Ekegren, et al., 2009). Most of the previous studies have used small numbers of physiotherapists and it has been noted that studies with broader cohorts of physiotherapists are also needed (Crossley, et al., 2011). Clinical experience is a factor that may influence the ability to visually assess movement (Whatman, et al., 2011b) and studies with larger numbers or raters are required to investigate this further. All previous studies we reviewed included substantial rater training prior to visual assessments. It is possible this additional training in some studies resulted in rating ability greater than could be expected in current clinical practice. To gain a picture of rating ability in current clinical practice a study of larger numbers of physiotherapists without additional training is needed. Additionally tests such as the drop jump occur far more quickly than SKB tests. It has been suggested that faster movements are more difficult to rate visually (Knudson, 2000), however there is little known about how this affects visual rating of dynamic alignment during these functional tests.

There is limited information on the validity of visual ratings of dynamic pelvis and knee alignment especially in comparison to kinematic data. Three-dimensional (3D) motion analysis is considered the gold standard for quantifying movement; however it is not applicable to general clinical use. Two-dimensional (2D) measurement techniques, such as the knee frontal plane projection angle (FPPA),

have also been used to quantify movement during lower extremity functional tests, but again this is not in common clinical use. Only two studies have compared visual ratings (knee position) to 3D motion analysis with varied results (Ageberg, et al., 2010; Ekegren, et al., 2009). No studies have compared ratings of young athletes or included comparisons of visual ratings of the pelvis. Three studies (Ageberg, et al., 2010; Stensrud, et al., 2011; Tofte KB, et al., 2011) have shown significant differences in 2D kinematics (p≤0.001) between visually rated groups but again this has not been investigated in young athletes. The ability of visual ratings to identify young athletes with substantially different pelvic and lower extremity kinematics as identified by 3D and 2D motion analysis techniques warrants further investigation.

Purpose

To investigate the ability of physiotherapists to visually rate knee and pelvic position in young athletes during lower extremity functional tests, specifically; (1) to investigate intra- and inter-rater reliability, (2) to investigate the validity of ratings and (3) to assess the influence of clinical experience and velocity of movement on rating ability.

Methods

Participants

Video was recorded of 23 (11 female) healthy young athletes (mean ±SD: 11 ±1 y, 153 ±10 cm, 44 ±8 kg) performing three lower extremity functional tests (SKB, Single leg SKB and Drop jump). All participants were part of a structured long term athlete development (LTAD) programme and competed in a variety of sports. The study was approved by a university ethics committee. All participants/parents received verbal and written information regarding the study and all participants gave assent prior to participation. All parents also provided written informed consent prior to testing. Sixty six New Zealand registered physiotherapists, recruited via website and email advertisements, agreed to visually rate the video recordings. All physiotherapists also provided written informed consent prior to the study. Twenty six of these physiotherapists repeated the visual ratings 3-4 weeks later. A period of two weeks has been suggested as the minimum period between viewing the videos

to avoid raters remembering initial ratings (Ekegren, et al., 2009). The number of athletes rated was primarily a resource decision based on the number of athletes in the LTAD programme.

Three lower extremity functional movement tests

All athletes performed three functional tests (see Table 6.1). The two small knee bend functional tests were chosen due to their common usage by physiotherapists in musculoskeletal practice and their ease of administration. The Drop Jump was included as it is the most commonly reported test in the literature for screening young athletes and for comparative purposes to the small knee bend tests. All athletes were given standardized verbal instructions prior to each test and the researcher demonstrated each test in a standardized manner. Practice for all tests was allowed until the researcher was happy the test was performed appropriately and consistently. Additional detail on the small knee bend functional tests and their reliability has been established previously (Whatman, et al., 2011a). All three functional tests were recorded from an anterior view on digital video (Panasonic, USA) sampling at a rate of 60 Hz. The video camera was positioned on a tripod in front of athletes and perpendicular to the frontal plane, at a height of 0.86 m and a distance of 3.7 m. All athletes were also instrumented with retro-reflective markers allowing a nine camera motion analysis system (Qualysis Medical AB, Sweden) sampling at 240 Hz to simultaneously collect three dimensional kinematic data.

Visual assessment procedure

Video clips of the young athletes performing the functional tests were produced using Windows™ Movie Maker. The order of the functional tests in each video clip was randomised. The video clips were placed on a university website to which all physiotherapists were given access. Standardised instructions on how to view the video clips and one practise video were available to all physiotherapists prior to them performing visual ratings. Video clips were set up such that each functional test, performed by each athlete, was looped three times (each repetition was thus identical). Pausing of the video clips between each test to record a rating was permitted. Pausing of the video clips during the performance of a functional test or repeating a clip was not permitted.

The visual assessment procedure was based on protocols we have used previously (Whatman, et al., 2011b) and those published by other authors (Ageberg, et al., 2010; Chmielewski, et al., 2007). For all three functional tests, physiotherapists visually assessed the position of the right patella relative to the 2nd toe during the loading/knee flexion phase of the test. A "yes" or "no" response was recorded to the question, "Does the patella move medial to the 2nd toe"? Additionally for the Single leg SKB, the position of the pelvis in the frontal plane was also rated. Rating the pelvis in single leg tests (but not double leg tests) is common clinically and has been reported in previous studies (Chmielewski, et al., 2007; Crossley, et al., 2011). Again a "yes" or "no" response was recorded to the question, "Does the pelvis remain neutral in the frontal plane"? No other specific guidelines/training was given and no examples of previously rated athletes were presented. All ratings were recorded directly on the website.

All video clips were also rated in the same manner by three expert musculoskeletal physiotherapists (all PhD or Masters qualified and all senior academics with an average of 15 years clinical experience) using video analysis software (Siliconcoach 7, New Zealand) which allowed slow motion, pause, replay and the overlay of horizontal and vertical lines. Ratings were completed independently and where there were any discrepancies the group reviewed the videos to reach a consensus as to the correct rating. These ratings formed the criterion against which the visual ratings of the other 66 physiotherapists were compared.

Table 6.1:   Description of the three lower extremity functional tests.

Functional test	Test description
Small knee bend (SKB)	Starting from a standing position, athletes performed a partial squat (hip and knee flexion) with the trunk maintained in an upright position while looking straight ahead. Athletes were instructed to continue the SKB until they reached maximum dorsiflexion without lifting their heels and then return to upright standing.
Single leg SKB	Standing on one leg at a time, with the contralateral hip in neutral and knee flexed to approximately 80°, athletes performed a SKB as described above.
Drop Jump	Athletes started on a 25 cm high box and dropped directly down off the box onto the force plate and immediately jumped vertically as high as possible.

Data processing

All visual rating data were downloaded into an excel spreadsheet and answers recoded as "1" or "2" for further analysis. All functional tests were tracked using the Qualysis motion capture software and exported to Visual 3D (C-Motion Inc, USA). In Visual 3D a rigid link model (pelvis, thigh, shank and foot) created from a static trial was assigned to all imported motion files to calculate kinematic data using established protocols (Whatman, et al., 2011a). All pelvis and lower limb kinematic data were exported to a customised Labview programme and processed to provide peak joint angles during the loading phase (knee flexion phase) of each functional test. The loading phase in the Drop Jump was from initial ground contact (determined by the force plate recordings based on the onset of the vertical ground-reaction force) to maximum knee flexion and for the small knee bend tests from minimum to maximum knee flexion. The two-dimensional (2D) video footage was also imported into Siliconcoach 7 for calculation of the peak knee FPPA and peak pelvic lateral tilt during the loading/knee flexion phase of each functional test. The knee FPPA was calculated as the angle formed by a line connecting the anterior superior iliac spine (ASIS) to the patella and a line connecting the patella to the midpoint of the ankle in a manner similar to that reported previously (Olson, et al., 2011; Willson & Davis, 2008b). The pelvic angle was calculated as the angle formed by a line connecting the ASIS's and a horizontal line. All 2D angles were measured by a single, experienced physiotherapist and the reliability of these measures was established in pilot testing (ICC≥0.98).

Statistical analyses

The level of intra-rater agreement for 26 physiotherapists and inter-rater agreement for all 66 physiotherapists and for groups based on experience was determined using overall percentage agreement (PA) and the first order agreement coefficient (AC1). For intra-rater agreement only two groups based on experience were calculated due to the low physiotherapist numbers. Overall percentage agreement was calculated as the number of observed agreements divided by the maximum number of possible agreements. The first order agreement coefficient (AC1) was calculated using a SAS macro published by Blood and Spratt (2007). The AC1 adjusts the overall percentage agreement for chance agreement giving an estimate of the true agreement. The calculation of the AC1 has been described in detail previously (Whatman, et al., 2011b). Gwet (2001) suggested the

AC1 can be interpreted in a similar manner to the kappa coefficient, based on a scale proposed by Landis and Koch (1977): 0.01-0.20=slight; 0.21-0.40=fair; 0.41-0.60=moderate; 0.61-0.80=good/ substantial; 0.81-1.0=almost perfect/excellent.

To assess the validity of ratings, sensitivity, specificity, positive predictive value, and negative predictive value were calculated by comparing the initial ratings of all 66 physiotherapists to the consensus ratings of the three expert physiotherapists. Means were calculated as an indication of group performance (as might be of interest in a large screening situation involving a number of physiotherapists) and ranges were calculated to indicate how an individual physiotherapist might perform in the clinic. These indicators of rater performance were derived from a 2 x 2 contingency table generated in SPSS version 18. Each statistic was then calculated in Microsoft Excel using standard formulas as reported by Glas, Lijmer, Prins, Bonsel, and Bossuyt (2003). In keeping with the suggestion of Ekegren et al., (2009) sensitivity (desired level set at 80%) was given priority over specificity (desired level set at 50%). Attaining high sensitivity ensures athletes with poor alignment are not classified as good. The main problem of low specificity is some athletes with good alignment might be rated as poor and given interventions that while low cost and not harmful are unnecessary. Validity was further assessed by comparing the peak 2D and 3D kinematic data between the groups based on the consensus, expert visual ratings [e.g. group rated "yes" for patella medial to 2nd toe (poor alignment) compared to group rated "no" (good alignment)]. The trials used to calculate the kinematics were the same as the trials rated visually by the experts. As suggested by Hopkins (2009) all conclusions were based on magnitude based inferences rather than null-hypothesis testing. Rather than the use of p values we have used inferences based on the uncertainty in the magnitude of the differences in mean kinematics between groups through the use of confidence intervals. Inferences were based on where the confidence interval lay in relation to threshold values for substantial differences and were qualified by probabilities using a scale published by Hopkins (2009).

To quantify the effect of experience and the velocity of a movement on rating ability we also calculated and assessed differences in the DOR. The DOR is a single indicator of rating performance that combines the strengths of sensitivity and specificity, is prevalence independent and recommended for comparing test performance (Glas, et al., 2003). The DOR has advantages over the

use of paired indicators (such as sensitivity and specificity) when comparing rating performance, particularly if one group/test does not outperform the other on both indicators (Glas, et al., 2003). The mean DOR of groups of physiotherapists with different levels of experience was compared for <5 versus 10-14 years and 5-9 versus >14 years. We decided on these comparisons as we wanted to maintain a minimum five year difference between groups. The log transformation of the DOR was used for all comparisons. Again the methods of Hopkins (2009) were used to quantify differences in rating performance (differences in the mean DOR between groups). To assess the influence of movement velocity we also assessed the difference in mean DOR for the Drop Jump compared to the SKB (based on ratings by all 66 physiotherapists).

Results

The 66 physiotherapists who participated in the study had a range of experience (<5 yrs n=15, 5-9 yrs n=16, 10-14yr n=15, >14 yrs n=20). Interpretation of the confidence intervals showed the effect of experience on rating accuracy with this sample size was clear (see Table 6.5). Sixty percent had a postgraduate qualification and all worked in the musculoskeletal area, regularly using the small knee bend test (or similar) to assess clients. In contrast only 19% reported regularly using the Drop Jump test. Sixty percent of the physiotherapists used the term small knee bend to identify the tests used in the study (we are aware of the many different names given to the types of functional tests we have used).

Rater agreement

Mean intra-rater agreement for all 26 physiotherapists who made repeat ratings was substantial for all tests (PA: 79% to 88%, AC1: 0.60 to 0.78) (see Table 6.2). Agreement for the small knee bend tests ranged from moderate to excellent and analysis of the 1st quartile revealed that 75% of physiotherapists achieved substantial intra-rater agreement or better (AC1: 1st quartile ≥ 0.60). Agreement for the Drop Jump ranged from slight to excellent with 50% of physiotherapists achieving substantial agreement (AC1: median ≥ 0.60). Mean agreement for the less experienced group of physiotherapists ranged from fair to substantial (PA: 76% to 86%, AC1: 0.56 to 0.78) while the more

experienced group achieved slightly better substantial to excellent agreement (PA: 82% to 90%, AC1: 0.65 to 0.81).

Inter-rater agreement was not as good as intra-rater agreement. For the small knee bend tests inter-rater agreement for all physiotherapists ranged from moderate to substantial (PA: 73% to 80%; AC1: 0.52 to 0.61) but agreement was only fair for the Drop Jump (PA: 67%, AC1: 0.37), (see Table 6.3). The group of physiotherapists with <5 years experience achieved the lowest inter-rater agreement for all tests (AC1: 0.32 to 0.47), but there is little difference in agreement between the other groups.

Validity

Mean sensitivity reached the target value (≥80%, see Table 6.4) for all tests (except the Drop Jump) and the majority of physiotherapist achieved this (1st quartile ≥ 80%). Mean specificity reached the target value (≥50%) for all tests and the majority of physiotherapists also achieved this (1st quartile ≥ 50%) except for knee ratings in the Single Leg SKB.

Five years more experience possibly improved rating accuracy in the SKB and Drop Jump (DOR: 1.6 to 2.8 times better, see Table 6.5). Having a postgraduate qualification did not improve rating accuracy for any test (DOR: 1.0 to 1.2 times better). Ratings of the SKB were almost certainly more accurate than those of the Drop Jump (DOR 4.9 times better, 90% CL, 3.4-7.0).

Based on the expert consensus ratings, athletes visually rated as having a patella medial to the 2nd toe alignment were likely to very likely to have increased peak 3D hip internal rotation and adduction in all tests (see Table 6.6). These athletes were also almost certain to have an increased peak knee FPPA in the small knee bend tests and very likely to have an increased peak knee FPPA in the Drop Jump. There was however no clear difference in the peak knee abduction angle between groups. Athletes rated as not maintaining a neutral pelvis in the frontal plane were almost certain to have increased lateral pelvic tilt as measured in 3D and 2D.

Table 6.2:   Intra-rater agreement.

	All physiotherapists (n=26)		< 10 yr experience (n=9)		> 14 yr experience (n=11)	
	PA	**AC1**	**PA**	**AC1**	**PA**	**AC1**
Small Knee Bend (K)	88 (74 to 100)	0.78 (0.49 to 1.00)	87 (83 to 91)	0.78 (0.67 to 0.87)	90 (78 to 100)	0.81 (0.60 to 1.00)
Drop Jump (K)	79 (57 to 96)	0.60 (0.14 to 0.92)	76 (57 to 91)	0.56 (0.20 to 0.83)	82 (61 to 96)	0.65 (0.22 to 0.91)
Single Leg SKB (K)	83 (70 to 96)	0.71 (0.41 to 0.95)	84 (70 to 96)	0.71 (0.41 to 0.95)	84 (74 to 91)	0.71 (0.49 to 0.87)
Single Leg SKB (P)	84 (70 to 100)	0.73 (0.48 to 1.00)	83 (74 to 100)	0.70 (0.48 to 1.00)	85 (70 to 96)	0.74 (0.52 to 0.93)

Mean (range), PA = Percent Agreement, AC1 = Agreement Coefficient 1, K=knee rating, P=pelvis rating.

Table 6.3:   Inter-rater agreement based on initial ratings.

	All physiotherapists (n=66)		< 5 yr experience (n=15)		5-9 yr experience (n=16)		10-14 yr experience (n=15)		> 14 yr experience (n=20)	
	PA	**AC1 (95% CL)**	**PA**	**AC1 (95% CL)**	**PA**	**AC1 (95% CL)**	**PA**	**AC1 (95% CL)**	**PA**	**AC1 (95% CL)**
Small Knee Bend (K)	80	0.61 (0.48 to 0.73)	71	0.42 (0.25 to 0.60)	82	0.68 (0.56 to 0.80)	83	0.67 (0.51 to 0.82)	81	0.63 (0.47 to 0.78)
Drop Jump (K)	67	0.34 (0.22 to 0.47)	66	0.32 (0.19 to 0.46)	69	0.33 (0.33 to 0.55)	68	0.37 (0.21 to 0.53)	68	0.36 (0.22 to 0.50)
Single Leg SKB (K)	73	0.52 (0.41 to 0.65)	70	0.44 (0.29 to 0.58)	72	0.50 (0.37 to 0.63)	72	0.50 (0.37 to 0.63)	75	0.55 (0.40 to 0.70)
Single Leg SKB (P)	73	0.52 (0.42 to 0.62)	70	0.47 (0.36 to 0.58)	75	0.53 (0.36 to 0.70)	74	0.54 (0.44 to 0.64)	74	0.53 (0.42 to 0.65)

PA = Percent Agreement, AC1 = Agreement Coefficient 1, K=knee rating, P=pelvis rating

Table 6.4: Accuracy of all physiotherapist initial visual ratings (percent) when compared to consensus expert ratings.

	Sensitivity	Specificity	+ve predictive value	-ve predictive value
Small Knee Bend (K)	88 (88 to 100)	85 (80 to 100)	81 (70 to 100)	94 (92 to 100)
Drop Jump (K)	70 (61 to 81)	79 (71 to 96)	89 (83 to 96)	56 (46 to 60)
Single Leg SKB (K)	95 (89 to 100)	54 (36 to 64)	62 (50 to 64)	96 (91 to 100)
Single Leg SKB (P)	86 (80 to 100)	66 (56 to 78)	81 (75 to 87)	80 (70 to 100)

Mean (interquartile range), K=knee rating, P=pelvis rating

Table 6.5: Comparison between rating ability (diagnostic odds ratio) of more and less experienced physiotherapists.

	<5 yrs (n=15) versus 10-14 yrs (n=15)	5-9 yrs (n=16) versus >14 yrs (n=20)
Small Knee Bend (Knee rating)	2.2 (0.8 to 5.9), 60%, possibly	1.6 (0.9 to 2.8), 26%, possibly
Drop Jump (Knee rating)	0.8 (0.5 to 1.4), 0%, most unlikely	2.8 (1.4 to 5.4), 79%, likely
Single Leg SKB (Knee rating)	1.2 (0.8 to 1.8), 3%, very unlikely	1.5 (0.9 to 2.3), 11%, unlikely
Single Leg SKB (Pelvis rating)	2.5 (1.3 to 4.5), 72%, possibly	1.0 (0.6 to 1.6), 1%, very unlikely

Difference in mean diagnostic odds ratio as a factor, (90% CL), likelihood (percentage and qualitative inference) the difference is clinically meaningful (more experienced physiotherapists ratings better by a factor >2).

Table 6.6:   Mean angle difference between groups rated yes (poor alignment) or no (good alignment) by the three experts.

	Small Knee Bend	Drop Jump	Single Leg SKB
Peak Hip Adduction (3D)	2.6 (0.4 to 4.7), 93%, likely	5.2 (2.2 to 8.1), 99%, very likely	6.0 (1.1 to 11.0), 94%, likely
Peak Hip Internal Rotation (3D)	6.5 (3.0 to 10.0), 99%, very likely	6.3 (2.4 to 10.2), 98%, very likely	4.0 (1.7 to 6.3), 99%, very likely
Peak Knee Abduction (3D)	1.1 (-2.2 to 4.4), 56%, possibly/possibly not	1.0 (-2.8 to 4.9), 51%, possibly/possibly not	2.3 (-1.3 to 5.9), 74%, possibly/possibly not
Peak Knee FPPA (2D)	16.1 (11.2 to 20.9), 100%, almost certainly	15.3 (5.3 to 25.3), 98%, very likely	20.7 (15.3 to 26), 100%, almost certainly
Peak Lateral Pelvic Tilt (3D)†			11.8 (8.6 to 14.9), 100%, almost certainly
Peak Lateral Pelvic Tilt (2D)†			8.5 (6.6 to 10.4), 100%, almost certainly

Difference in means (degrees) between groups (90% CL), likelihood (percentage and qualitative inference) the difference is clinically meaningful (> Cohen 0.2), FPPA=Frontal plane projection angle. † Not rated in the Small Knee Bend or Drop Jump

Discussion

The identification of poor frontal plane pelvis, hip and knee control is considered important. The validation of simple, easily administered, inexpensive clinical tests to achieve this is needed. This study investigated the reliability and validity of physiotherapist visual rating of dynamic knee and pelvic position in young athletes. The influence of experience, having a postgraduate qualification and the velocity of the movement was also examined. Our rating approach was based on our earlier work and that of others suggesting keys to achieving accurate ratings included simple dichotomous scales and the use of anatomical landmarks where possible (Ageberg, et al., 2010; Chmielewski, et al., 2007; Ekegren, et al., 2009; Whatman, et al., 2011b). In an effort to capture current ability in clinical practice we recruited a broad range of physiotherapists and our rating guidelines did not involve detailed instructions or rater training.

Mean intra-rater agreement for the knee and pelvis was substantial for all tests and most physiotherapists achieved this level of agreement. Agreement was not improved by experience but knee rating was better for the slower, small knee bend tests than for the faster Drop Jump. The level of intra-rater agreement was similar to that reported for ratings of the pelvis (77-82%) and knee (80-84%), during small knee bend tests in adults (Whatman, et al., 2011b). Agreement for knee ratings was slightly better than that reported by Örtqvist et al., (2011) for a single physiotherapist rating knee position, in children/adolescents aged 9-16 years, during a single limb mini squat (PA=76%). Our knee ratings were also similar to that reported previously by Ekegren et al., (2009) for visual ratings of medio-lateral knee position in adolescent female soccer players performing a drop jump (PA=88 to 90%). Ekegren et al., (2009) used a similar, simple rating method based on patella position but gave raters (three experienced physiotherapists) detailed instructions and training.

The level of inter-rater agreement was moderate to substantial for the small knee bend tests but only fair for the faster Drop Jump. The least experienced physiotherapists (<5 yrs) achieved slightly lower agreement for all tests than the other physiotherapists. The level of inter-rater agreement for knee ratings was similar to that reported for two physiotherapists (PA=79%) in the study by Örtqvist et al., (2011) and similar to the agreement reached in ratings of the pelvis (PA=70%) and knee (PA=78%)

for adults performing small knee tests (Whatman, et al., 2011b). Ageberg et al., (2010) reported higher agreement (PA=96%) for two physiotherapists in a study rating medio-lateral knee position during a single limb mini squat, however there was thorough rater training. Higher inter-rater agreement (κ = 0.77 to 0.80) was also reported in the study by Ekegren et al., (2009) that also involved substantial rater training. We consider the inter-rater agreement for the small knee bend tests in our study acceptable, especially when considered alongside the high sensitivity achieved. High sensitivity suggests rating disagreements will mainly result in athletes with good alignment being rated by some physiotherapists as poor and given harmless, although unnecessary intervention. It is possible agreement could be improved with rater training and this may be desirable in some circumstances. Additionally although our inter-rater reliability was slightly lower than previously reported we looked at a much larger group of physiotherapists which likely better reflects broader clinical practice.

Despite their common use the validity of visual ratings of the knee and pelvis has not been clearly established, especially in young athletes. Using the expert consensus ratings as our criterion the group (and the majority of physiotherapists individually) achieved acceptable sensitivity for both small knee bend tests but not the Drop Jump. High sensitivity ensures athletes who actually have poor alignment will be detected by the physiotherapists and thus would be given the appropriate intervention to correct their technique. Additionally high sensitivity means relatively few athletes will be incorrectly rated as having good alignment and attests to the value of these visual ratings for ruling out truly poor alignment. In keeping with our findings sensitivity has previously been reported as inadequate (67-87%) for ratings of knee position during the Drop Jump (Ekegren, et al., 2009).

Specificity was also high for the SKB attesting to the value of this test for correctly identifying athletes with good alignment and thus not prescribing unnecessary intervention. Although not as good, specificity was still acceptable for the Drop Jump and Single Leg SKB although 20% to 46% of athletes would be rated as having poor alignment when in fact they don't. Given the low cost, harmless nature of neuromuscular interventions designed to improve alignment we don't consider this a major concern. Others however have raised the question of the undesirable stigma of being classified as having poor movement and the possible increased fear of injury which may need further

consideration (Ekegren, et al., 2009). With regards to the Single Leg SKB test we noted substantial medio-lateral movement (oscillation) of the patella in several athletes. This could be because this slow, controlled single leg movement places high demands on strength and balance in these young athletes. It may be that the ratings of some physiotherapists have been influenced by the excursion of the patella rather than the absolute position relative to the 2nd toe and this lowered specificity. The addition of further simple instructions could likely alleviate this issue.

As indicated by differences in the DOR there was evidence that increased experience improved the accuracy of visual ratings of the knee in the SKB and Drop Jump and the pelvis in the Single Leg SKB. Only the most experienced group achieved more accurate Drop Jump ratings possibly reflecting the greater difficulty in rating this faster movement. The differences in the DOR represent approximately a 10% improvement in sensitivity if specificity remains unchanged and vice-versa. As most physiotherapists reached adequate sensitivity irrespective of experience, the major practical implications of this would likely be less wasted resources with improved specificity resulting in less athletes receiving unnecessary intervention. The influence of experience was however not consistent across all ratings. This may reflect rating movement from video was new to the majority of physiotherapists and experience with this method was similar across the group. Previous studies have suggested experience improves rater agreement (Brunnekreef et al., 2005; Whatman et al., 2011b) and it seems possible it may also improve the accuracy of ratings. Alternatively success with visual ratings may be more a consequence of the perceptual style of the observer (Morrison, 2000). Having a postgraduate qualification did not improve rating accuracy but as anticipated the SKB was more accurately rated than the Drop Jump. This confirms the view of Knudson (2000) who suggested velocity of movement was a key factor in visual rating accuracy. If sufficient information can be gained from the SKB it is the more valid clinical tool. The use of video slow motion may need to be considered to achieve more accurate ratings of the Drop Jump; however this has resource implications for clinical practice. We acknowledge the Drop Jump was not a test commonly used by our group of physiotherapists and this may have also contributed to the poorer rating performance.

Although predictive values are highly dependent on prevalence in the sample it is suggested they may be easier to interpret for a clinician than sensitivity and specificity (Fritz & Wainner, 2001). They relate

to the way the tests are used in clinical decision making (i.e. given a rating, poor or good alignment, what is the probability the rating is correct?). The high negative predictive values for the small knee bend tests (especially the knee ratings) indicate a rating of good alignment is very likely correct. This is important from a screening point of few as it means that few athletes with poor knee alignment have been incorrectly classified. The negative predictive value for the Drop Jump was unacceptably high with approximately half the athletes with poor knee alignment incorrectly rated. The positive predictive values seem acceptable other than for the knee rating in the Single Leg SKB. Approximately 40% of athletes rated as poor may have actually had good alignment resulting in unnecessary interventions which although harmless may impact on resources.

The group visually rated by the expert panel as having a patella medial to the 2nd toe position displayed greater knee FPPA, hip adduction and hip internal rotation in all tests, but no difference in knee abduction angle. This suggests the appearance of the patella medial to the 2nd toe was mainly due to altered hip kinematics. This is similar to the findings of Ageberg et al., (2010) who also reported increased hip internal rotation but not knee valgus in a group visually rated as knee medial to the foot during a single-limb mini squat. Our results attest to the validity of ratings for identifying athletes with kinematics that place them at higher risk of injury. Increased hip internal rotation and adduction during functional tasks has been associated with patellofemoral pain syndrome (PFPS) (Boling, et al., 2009; Souza & Powers, 2009; Willson & Davis, 2008a). The association between hip control and PFPS is further supported by two studies that have reported deficits in hip strength and neuromuscular control in participants with PFPS (Cowan, et al., 2008; Ireland, Willson, Ballantyne, & Davis, 2003). Hip weakness may predispose female athletes to increased femoral adduction and internal rotation and thus greater risk of lower extremity injury (Leetun, Ireland, Willson, Ballantyne, & Davis, 2004). An increase in the knee FPPA has also been associated with PFPS (Willson & Davis, 2008b). Additionally the group rated as not maintaining a neutral pelvis in the frontal plane showed increased lateral pelvic tilt. Assessment of pelvis position has been shown to be useful in identifying hip muscle dysfunction (Crossley, et al., 2011).

The major limitations of this study were the rating of healthy athletes and that ratings were made via video, from an anterior view only. Ratings of injured athletes may have given us an indication of

reliability and validity across a wider range of movement patterns and this is an area that needs further investigation. We emphasise that the rating ability reported in this study only pertains to healthy young athletes and may not be representative of an injured population. Further studies are required to assess the use of these tests in identifying readiness for return to sport and progress during rehabilitation with injured athletes. However, visual rating of dynamic alignment is used clinically as part of a screening process for risk of injury and our results are likely to give an indication of reliability and validity in these circumstances.

In this study only video from an anterior view was utilised as this appears common in the literature and also clinically for rating frontal plane control. It is possible that this 2D projection of a movement pattern may not be an adequate representation of movements in some planes or an ideal simulation of what is conducted clinically where movement is mostly rated live in three dimensions. The decision to use video rather than live rating was partly pragmatic (each athletes testing session took an hour) and because we wanted all physiotherapists to see exactly the same movement. We also acknowledge that the presence of anatomical markers may have improved ratings. As the markers were small we feel this effect is likely to have been negligible, however, further research without markers is needed to quantify the effect of markers on ratings.

Conclusion

A broad range of physiotherapists', without the need for additional training, were able to reliably rate dynamic knee and pelvic position in young athletes. Sensitivity and specificity was also acceptable for all ratings of the small knee bend tests when compared to the video assisted ratings of three experts. The faster Drop Jump achieved acceptable specificity but lacked sensitivity and was not rated as accurately or reliably as the slower SKB test. Increased experience may improve rating accuracy but this was not consistent across all ratings. For all tests the ratings of the experts were able to differentiate athletes with different hip and pelvis kinematics that might place them at increased risk of injury. Visual rating by physiotherapists should be able to reliably and accurately identify young athletes with poor frontal plane pelvis and knee control that needs correction.

CHAPTER 7: THE ASSOCIATIONS BETWEEN 2D ALIGNMENT AND 3D KINEMATICS DURING LOWER EXTREMITY FUNCTIONAL TESTS IN YOUNG ATHLETES

Overview

Purpose: To investigate the associations between two-dimensional (2D) measures of alignment [knee frontal plane projection angle (FPPA) and lateral pelvic tilt] and three-dimensional (3D) frontal and transverse plane pelvis, hip and knee kinematics in young athletes during lower extremity functional tests. *Methods:* Pelvis and lower extremity 3D kinematics at peak knee flexion were quantified in 23 uninjured young athletes (11 ±1 years) during three lower extremity functional tests (Small Knee Bend [SKB], Single Leg SKB and Drop Jump). A nine camera motion analysis system captured all tests from retro-reflective body markers and Visual3D and Labview were used to process all 3D data. Lateral pelvic tilt and knee FPPA from 2D video were calculated using SiliconCoach7. Magnitudes of Pearson correlation coefficients were used to describe associations between the 3D kinematics and the 2D measures of alignment. *Results:* There were moderate (r = 0.33 to 0.47) associations between knee FPPA and hip joint kinematics and small to moderate (r = 0.21 to 0.40) associations with knee joint kinematics. There were large to very large (r = 0.71 to 0.88) associations between knee FPPA and femoral and tibial segment frontal plane positions. There was a very large association between lateral pelvic tilt measured in 2D and 3D (r = 0.71, 90% CL, 0.56 to 0.81). *Conclusions:* Simple 2D measures of lower extremity alignment may be useful for identifying kinematics that should be suspected of placing young athletes at risk of injury; however they should not be used to quantify 3D joint angles.

Background

Assessment of lower extremity dynamic alignment is recommended when screening young athletes for injury risk, during rehabilitation or for performance enhancement (Örtqvist, et al., 2011). Due to the cost and time involved in a full biomechanical analysis, assessment in the clinic is usually performed via visual observation of functional tests. There is evidence visual assessment of pelvic position and

knee position relative to the foot is a useful clinical tool with acceptable reliability (Ageberg, et al., 2010; Whatman, et al., 2011b). However studies have shown that visual assessment differentiates participants most clearly on two-dimensional (2D) frontal plane kinematics (Ageberg, et al., 2010; Tofte, et al., 2011). A limitation of these studies comparing visual ratings to 2D kinematic analysis is the uncertainty in the link between these 2D measures and three-dimensional (3D) kinematic measures that have been linked to injury. Excessive lower extremity valgus (based on 3D kinematics) is thought to place athletes at increased risk of injury such as noncontact anterior cruciate ligament tear and patellofemoral dysfunction (Hewett, et al., 2005; Powers, 2003). A lack of pelvis and hip control is considered a key factor in the development of this valgus alignment (Powers, 2010). Thus the association between 2D and 3D kinematics is important and has implications for the validity of visual assessment of alignment. Additionally 3D motion analysis is not practical in most clinical settings while 2D measurement techniques using simple, portable, inexpensive equipment could be a useful clinical alternative and/or complement to visual assessment in some situations. Further studies are required to clarify the association between 2D measures of pelvic and lower extremity alignment and 3D kinematics in young athletes.

Poor dynamic alignment is a combination of poor frontal and/or transverse plane control of the pelvis, hip, knee and foot (Earl, et al., 2005; Powers, 2003; Sahrmann, 2002; Willson & Davis, 2009). The most commonly reported measure of this alignment in 2D appears to be the knee frontal plane projection angle (FPPA) (Herrington, 2009; Olson, et al., 2011; Willson & Davis, 2008b). It is also common clinically to measure the lateral tilt of the pelvis. The drop jump is the most commonly reported lower extremity functional test for assessing lower limb dynamic alignment in young athletes (Ford, et al., 2007). However the use of other hip and knee flexion tests is common clinically including the small knee bend (otherwise known as the partial squat). Single leg tests are also used and the single-leg mini squat has been recommended for use in young athletes (Örtqvist, et al., 2011). The reliability of kinematics during the small knee bend test and the drop jump (ICC=0.62 to 0.98) has been reported previously (Whatman, et al., 2012a).

The 2D method of measuring knee FPPA has been reported as associated (r = 0.50-0.60) with 3D calculated knee valgus during side jump, side step and shuttle run tasks in adult male college

basketball players (McLean, et al., 2005). Willson and Davis (2008b) also reported a moderate association (r = 0.32-0.48) between 2D knee FPPA and 3D hip adduction and knee external rotation, and a moderate association (r = 0.49-0.61) between knee FPPA and 3D femoral and tibial rotations, during a single leg squat in females with and without patellofemoral pain. In contrast Ekegren et al., (2009) reported 3D joint angles in a drop jump were not associated with corresponding 2D angles measured from video. Olson et al., (2011) in a study of adult females concluded that exercises to improve lower extremity alignment may improve knee FPPA but this may not be related to specific changes in 3D joint kinematics. There have not been any studies of the associations between 2D measures of alignment (knee FPPA and lateral pelvic tilt) and 3D frontal and transverse plane pelvis, hip and knee kinematics in young athletes during common lower extremity functional tests.

Purpose

To investigate the associations between 2D measures of alignment (knee FPPA and lateral pelvic tilt) and 3D frontal and transverse plane pelvis, hip and knee kinematics in young athletes during three common lower extremity functional tests (Small Knee Bend, Single Leg SKB and Drop Jump).

Methods

Participants

Twenty three young athletes (mean ±SD: 11 ±1 y, 153 ±10 cm, 44 ±8 kg) with no musculoskeletal problems volunteered for this study. All athletes were part of a structured long term athlete development programme and competed in a variety of sports. The study was approved by the university ethics committee. All participants/parents received verbal and written information regarding the study and all athletes gave assent prior to participation. All parents also provided written informed consent prior to testing.

Instrumentation

A nine camera motion analysis system (Qualysis Medical AB, Sweden) sampling at 240 Hz collected kinematic data. Athletes' pelvis and both legs had retro-reflective markers (19 mm diameter) secured

to anatomical locations (sacrum, bilateral ASIS's, iliac crests, greater trochanters, medial and lateral femoral epicondyles, mid-patella, medial and lateral malleoli) by an experienced musculoskeletal physiotherapist. Four cluster marker sets were attached to the thigh and shank of each leg. The anatomical markers were used for construction of a skeletal model using Visual 3D (C-Motion Inc, USA).

Testing protocol

All twenty three athletes each attended the motion analysis laboratory on one occasion. Following instrumentation of the retro-reflective markers a static standing trial was collected. The order of the three functional tests was randomized among athletes and all athletes performed three trials of each test. For all tests athletes were given standardized verbal instructions prior to each test (Table 7.1) and the researcher demonstrated each test. Further details on these particular tests and their reliability have been published previously (Whatman, et al., 2011a). The Small Knee Bend (SKB), and the Single Leg SKB on each leg, were performed by all athletes with the range of motion determined by their maximum ankle dorsiflexion range of motion. For the Drop Jump athletes started on a 25 cm high box and dropped directly down off the box onto the force plate and immediately jumped vertically as high as possible. All three functional tests were recorded from an anterior view on digital video (Panasonic, USA) sampling at a rate of 60 Hz and simultaneously by the 3D system. The video camera was positioned on a tripod in front of athletes and perpendicular to the frontal plane, at a height of 0.86 m and a distance of 3.7 m. All 3D kinematics and 2D measures of alignment were calculated from the same trial of each test.

Table 7.1:   Description of the small knee bend functional tests used in the study.

Functional test	Test description
Small Knee Bend (SKB)	Starting from a standing position, athletes performed a partial squat (hip and knee flexion) with the trunk maintained in an upright position while looking straight ahead. Athletes were instructed to continue the SKB until they reached maximum dorsiflexion without lifting their heels and then return to upright standing.
Single Leg SKB	Standing on one leg, with the contralateral hip in neutral and knee flexed to approximately 80°, athletes performed a SKB as described above.

Data processing

All static and motion (functional test) trials were tracked using the Qualysis motion capture software and exported to Visual 3D (C-Motion Inc, USA). In Visual 3D the rigid link model (pelvis, thigh, shank and foot) created from the static file was assigned to all imported motion files to calculate kinematic data. Data were filtered with a second-order Butterworth bidirectional low-pass filter with a cut-off frequency of 6 Hz. The Cardan sequence y-x-z was used for the calculation of joint angles which was equivalent to flexion/extension, abduction/adduction, axial rotation and in this case equivalent to the Joint Coordinate System described by Grood and Suntay (1983). The only exception to this was for calculation of pelvic angle with respect to the laboratory where we used the sequence z-x-y (axial rotation, obliquity, tilt) which is recommended in the Visual 3D documentation based on the suggestions of Baker (2001).

All pelvis and lower limb kinematic data were exported to a customised Labview programme and processed to provide joint and segment angles at maximum knee flexion. The two-dimensional (2D) video footage was imported into SiliconCoach7 (New Zealand) for calculation of the knee FPPA and 2D lateral pelvic tilt also at maximum knee flexion. The knee FPPA was calculated as the angle formed by a line connecting the ASIS to the patella and a line connecting the patella to the midpoint of the ankle malleoli in a manner similar to that reported previously (Olson, et al., 2011; Willson & Davis, 2008b) (see Figure 7.1). The angle was negative if the patella was medial to a line from the ASIS to

the midpoint of the ankle malleoli. The 2D pelvic angle was calculated as the angle formed by a line connecting the ASIS's and a horizontal laboratory aligned line (see Figure 7.1). All 2D angles were measured by a single, experienced physiotherapist and the reliability of these measures was established in pilot testing (ICC ≥ 0.98).

Figure 7.1:   Example of (A) Knee frontal plane projection angle and (B) 2D lateral pelvic tilt, during a Single Leg SKB.

Statistical analyses

Pearson correlation coefficients were calculated to assess the magnitudes of the associations between the 2D measures of alignment and 3D pelvis, hip and knee kinematics. Correlations were based on 46 kinematic measures (right and left legs from all 23 athletes). The magnitudes of these correlations were described as trivial (0.0-0.1), small (0.1-0.3), moderate (0.3-0.5), large (0.5-0.7), very large (0.7-0.9), or extremely large (0.9-1.0) (Hopkins, et al., 2009).

Results

There were moderate (r = 0.33 to 0.47) associations between knee FPPA and hip joint kinematics and small to moderate (0.21 to 0.40) associations with knee joint kinematics (see Table 7.3). There were large to very large (0.71 to 0.88) associations between knee FPPA and femoral and tibial segment frontal plane positions. Knee FPPA was also moderately associated with pelvic segment frontal plane position, but it was not associated with pelvic transverse plane position. There was a very large association between lateral pelvic tilt measured in 2D and 3D pelvic lateral tilt (r = 0.71, 90% CL, 0.56 to 0.81).

Table 7.2:   Angle and segment positions (°) at maximum knee flexion (mean both legs ± SD) for 23 healthy young athletes.

	Small Knee Bend	Drop Jump	Single Leg SKB
Hip Frontal*	-0.9 ±5.9	2.2 ±5.3	14 ±6.0
Hip Transverse†	1.7 ±6.0	-0.2 ±6.3	2.5 ±6.4
Knee Frontal‡	3.4 ±5.8	-0.9 ±5.9	6.5 ±5.2
Knee Transverse†	-1.7 ±5.2	-0.5 ±7.3	-4.2 ±6.6
Femur Frontal*	-1.1 ±4.7	1.9 ±4.8	11.2 ±4.9
Femur Transverse†	1.7 ±7.1	-0.3 ±8.3	4.9 ±8.4
Tibia Frontal‡	2.5 ±3.9	0.9 ±5.1	3.9 ±4.2
Tibia Transverse†	-3.3 ±6.0	1.6 ±8.1	9.9 ±13.8
Pelvis Frontal	-	-	-7.7 ±9.0
Pelvis Transverse	-	-	-1.9 ±7.1
Knee FPPA	3.4 ±10.1	-3.9 ±10.1	-8.4 ±10.6
2D Lateral pelvic tilt	-	-	4.2 ±3.7

* +ve = adduction, † +ve = internal rotation, ‡ +ve = abduction, - pelvic angle not measured in the small knee bend or drop jump, FPPA = Frontal plane projection angle.

Table 7.3:   Associations between knee frontal plane projection angle and 3D joint/segment transverse and frontal plane kinematics expressed as Pearson correlation coefficients (90% CL).

	Small Knee Bend	Drop Jump	Single Leg SKB
Hip Frontal	0.37 (0.14 to 0.56)	0.47 (0.25 to 0.64)	0.43 (0.21 to 0.61)
Hip Transverse	0.46 (0.24 to 0.63)	0.35 (0.11 to 0.55)	0.33 (0.09 to 0.53)
Knee Frontal	0.21 (-0.04 to 0.43)	0.29 (0.05 to 0.50)	0.23 (-0.02 to 0.45)
Knee Transverse	0.45 (0.23 to 0.63)	0.33 (0.09 to 0.53)	0.40 (0.17 to 0.59)
Femoral Frontal	0.79 (0.68 to 0.87)	0.71 (0.56 to 0.81)	0.77 (0.65 to 0.85)
Femoral Transverse	0.42 (0.19 to 0.60)	0.51 (0.03 to 0.67)	0.42 (0.19 to 0.60)
Tibial Frontal	0.88 (0.81 to 0.93)	0.81 (0.70 to 0.88)	0.85 (0.76 to 0.91)
Tibial Transverse	0.07 (-0.18 to 0.31)	0.08 (-0.17 to 0.32)	0.13 (-0.12 to 0.36)
Pelvic Frontal	-	-	0.31 (0.07 to 0.52)
Pelvic Transverse	-	-	0.06 (-0.19 to 0.30)

- pelvic angle not measured in the small knee bend or drop jump.

Discussion

The purpose of this study was to investigate the associations between 2D measures of alignment (knee FPPA and lateral pelvic tilt) and 3D frontal and transverse plane pelvis, hip and knee kinematics in young athletes during three common lower extremity functional tests (Small Knee Bend, Single Leg SKB and Drop Jump). Young athletes with poor dynamic alignment that increases their risk of injury need to be identified. Visual assessment and/or simple, portable, inexpensive measures of 2D alignment may be the most appropriate for clinical use.

The knee FPPA was not strongly correlated to the 3D hip and knee joint angles. There were moderate associations of the knee FPPA with hip frontal (adduction) and hip and knee transverse plane angles but only small correlations with knee frontal plane (abduction) angles. These findings are in keeping with those of Willson and Davis (2008b) who also found the weakest correlation was with knee joint valgus (r = 0.21) in a group of adult females with and without patellofemoral pain during a single leg squat. Willson and Davis (2008b) suggested this may be because there are multiple combinations of pelvic, hip and knee rotations that might contribute to a larger negative knee FPPA and this also appears to be the case in our young athlete population. As such the magnitude of 3D lower extremity transverse and frontal plane joint rotations in young athletes cannot be accurately determined by the magnitude of the 2D knee FPPA.

Of the 3D joint rotations, the hip joint showed the strongest association with knee FPPA. This may be important to clinicians given the recent focus on hip function and lower extremity injury. A review by Reiman et al., (2009) concluded there was mounting evidence that hip weakness contributes to knee injuries across all ages and Powers (2010) made a strong case for the link between abnormal hip kinematics and injury. Simple clinical assessment of the knee FPPA visually or assisted by video may help identify those young athletes in need of interventions designed to improve hip kinematics in weight bearing.

The 2D knee FPPA was most strongly associated with 3D femoral and tibial segment frontal plane position. Larger negative knee FPPA's (patella positioned more medially) were associated with

121

greater femoral adduction and tibial abduction. This is again in keeping with the findings of Willson and Davis (2008b) who also reported these as the strongest associations (r = 0.49 to 0.61). In the current study knee FPPA also had a moderate to large association with femoral transverse plane segment position (larger negative knee FPPA's were associated with increased femoral internal rotation). The smaller associations between the FPPA and 3D joint rotations (compared to segment rotations) may be due to relative rotations of two segments in the laboratory coordinate system. For example, simultaneous frontal plane motion of the femur and pelvis may result in little change in the hip joint frontal plane angle. Excessive adduction/internal rotation of the femur and/or abduction of the tibia are suggested to contribute to overuse injuries such as patellofemoral pain (Powers, 2010). These suggestions however appear predominantly based on theory considering the relative position of two segments at a joint (joint rotation) rather than the individual segment positions. Nevertheless the knee FPPA may be useful in identifying young athletes with femoral and tibial alignment that places them at increased risk of injury. Given the practical and time benefits for clinical use it may be sufficient to assess this knee FPPA using visual rating of functional tests such as the small knee bend or mini squat. Visual rating of alignment in young athletes during these functional tests has been shown to differentiate athletes based on knee FPPA (Whatman, et al., 2012b).

There was a strong association between the 2D measure of lateral pelvic tilt and the magnitude of 3D pelvic rotation in the frontal plane. Assessment of weight bearing pelvic alignment is common clinically as it is thought to give a further indication of hip muscle function as outlined by Crossley et al., (2011). Again this assessment of pelvis alignment is commonly made via visual observation in the clinic and this has been shown to be reliable in young athletes (Whatman, et al., 2012b). Visual assessment and/or 2D measures of pelvic alignment should therefore be useful in identifying young athletes with a 3D lateral pelvic tilt that is suggestive or poor hip muscle function.

A limitation of this study was the lack of exact synchronisation between the 2D and 3D measures. The 3D data were collected at a much higher sampling rate than the 2D and thus the 2D image used to measure the knee FPPA and pelvic tilt may not have occurred at the exact time point of the 3D data. We have also only taken measures at the single time point of maximum knee flexion and as such our findings do not indicate how 2D and 3D measures compare throughout the tests (e.g. at

different degrees of knee flexion or for joint excursions). Maximum knee flexion (mean ±SD: SKB = 90 ±17º, Single SKB = 84 ±10º, Drop Jump = 90 ±11º) was chosen as this is where maximum frontal plane deviation is likely to be seen clinically and this was also the easiest point to identify visually to synchronise the 2D and 3D angles. Additionally, although the total number of limbs we reported on is in keeping with similar previous studies (Willson & Davis, 2008b), we had a relatively small sample size of only 23 young athletes. This could lead to greater uncertainty in the true associations in the population. Precision for a correlation is thought to be adequate when the uncertainty in the estimate (represented by its confidence interval) does not span more than two qualitative magnitude thresholds (Hopkins & Manly, 1989). Our confidence intervals show that we appear to have adequate precision for the majority of correlations. The lower correlations still provide useful information but they must be interpreted with more caution.

Conclusion

Two-dimensional measures of alignment during lower extremity functional tests demonstrated small to moderate associations with 3D hip and knee kinematics but should not be considered a substitute for quantification of 3D joint rotations. The 2D kinematic measures may provide a useful clinical indication of young athletes at risk of injury and warranting appropriate intervention.

CHAPTER 8: VISUAL AND KINEMATIC ASSESSMENTS OF DYNAMIC ALIGNMENT DURING LOWER EXTREMITY FUNCTIONAL SCREENING TESTS: A REVIEW OF THE LITERATURE

Overview

Background: Assessment of lower extremity dynamic alignment/movement quality during functional screening tests, with evaluation of multiple joints in a chain rather than assessment of individual joints alone is an important part of musculoskeletal assessment. In clinical practice assessment is predominantly via visual observation, as clinicians do not have access to equipment or time required for complex biomechanical analysis. Clinicians must consider a number of factors including the reliability and validity of visual assessment and the reliability of kinematics during the tests. *Aim:* To investigate the evidence for visual and kinematic assessments of movement quality/dynamic alignment during lower extremity functional screening tests. *Methods:* Three relevant books and 33 journal articles were reviewed after searches of electronic databases using keywords 'lower extremity', 'alignment', 'knee', 'functional tests', 'squat', 'partial squat', 'single leg squat', 'drop jump', 'small knee bend', 'lunge', 'step-down', 'kinematics', 'reliability', 'validity', 'visual assessment' and 'visual rating'. *Findings:* Reliability of lower extremity joint kinematics during functional tests ranged from fair to excellent (ICC = 0.46 to 1.0). Based on studies using 3D motion analysis, with a range of age groups and a variety of timeframes, the best evidence of reliable kinematics exists for the single leg squat, small knee bend (single or double leg) and drop jump. Other studies were limited by a lack of three dimensional (3D) motion analyses, failing to report reliability both within-day and between-days, a variety of test protocols (some less suitable for clinical use and several with no reports of reliability), few reports of young athletes performing tests other than the drop jump, few studies considering the trunk and pelvis, studies only assessing the dominant side and a lack of absolute measures of reliability in degrees. Intra-rater agreement with visual assessment ranged from slight to excellent (AC1 = 0.01 to 0.96) and inter-rater agreement also ranged from slight to excellent (κ = 0.00 to 0.92). Visual assessment can differentiate groups with kinematics and hip muscle function that may place them at increased risk of injury. Clinician's visual ratings agreed sufficiently with those of an

expert consensus panel. Reliability and validity of visual assessment was generally improved with clinical experience, rater training, simple dichotomous rating methods and slower velocity tests. *Conclusion:* The most valid assessment of lower extremity dynamic alignment, in current clinical practice, will be achieved with the use of a single leg squat (SLS) or single of double leg small knee bend (SKB), rating knee position relative to the foot on a dichotomous scale. If using more complex rating methods provide example ratings and/or additional rater training. Further development of other tests and rating methods focusing on protocols with maximum clinical utility is needed.

Introduction

Visual assessment of movement quality/dynamic alignment is commonly used in clinical practice as a key tool in the prevention and management of musculoskeletal injuries. Sahrmann (2002) has promoted the evaluation of movement quality for many years for the prevention of musculoskeletal pain. Visual screening of quality of fundamental movements in all populations is now promoted with the goal of injury reduction, improved performance and enhanced quality of life (Cook, Burton, & Hoogenboom, 2006). A lack of lower extremity control in the frontal and transverse planes is suggested as a key risk factor in a number of injuries such as patellofemoral pain syndrome (Powers, 2003) and iliotibial band syndrome (Reiman, et al., 2009). In a review, Reiman et al., (2009) concluded there was mounting evidence that faulty lower extremity movement patterns contribute to knee injuries across all ages. Authors in another recent review suggested the inclusion of targeted proximal interventions (e.g. gluteal strengthening) may be key to reducing symptoms of anterior knee pain (Collins, Bisset, Crossley, & Vicenzino, 2012). A lack of dynamic lumbopelvic and knee control is also considered a risk factor for hamstring muscle injuries (Sole, Milosavljevic, Sullivan, & Nicholson, 2008). Additionally a link has been shown between kinematics during lower extremity functional tests and actual function (running and jumping) (Whatman, Hing, & Hume, 2011a; Willson & Davis, 2008a). Clinicians (physiotherapists, sports physicians, trainers) using lower extremity functional screening tests to visually assess movement quality/dynamic alignment must consider a number of factors: (i) the range of tests available; (ii) the reliability of kinematics during the various tests and (iii) the reliability and validity of visual assessments. This paper reviews the evidence for these issues and

provides recommendations for the use of lower extremity functional screening tests and future research.

Methods

A narrative review to investigate the evidence for the visual and kinematic assessments of movement quality/dynamic alignment during lower extremity functional screening tests was performed. Electronic data bases including SportsDiscus, Scopus, Medline, AMED, CINHAL, PEDro and the Cochrane Library were searched using keywords 'lower extremity', 'alignment', 'knee', 'functional tests', 'squat', 'partial squat', 'single leg squat', 'drop jump', 'small knee bend', 'lunge', 'step-down', 'kinematics', 'reliability', 'validity', 'visual assessment' and 'visual rating'. Exclusion criteria included articles that were (i) unavailable in English and not previously referred to by other sources; (ii) not specific to visual and kinematic assessment of dynamic alignment during lower extremity functional screening tests and did not add knowledge to the manuscript. Additional supportive articles were sought through article reference lists. Thirty six references were retained after determining the relevance of the information to the aim of the paper.

Findings

Common lower extremity functional screening tests

A range of lower extremity functional tests have been reported in the literature for assessing movement quality/dynamic alignment (see Table 8.1). The focus of all these tests is the assessment of frontal and/or transverse plane control of all or some of the trunk, pelvis, hip, knee and foot. Assessment of knee position relative to the foot is also common as it has been suggested that during hip and knee flexion the path of the knee should follow the longitudinal axis of the foot (the second toe) (Sahrmann, 2002). The most common lower extremity functional test is the single leg/limb squat (SLS) (Zeller, et al., 2003), alternatively called a unilateral squat (Weir, et al., 2010) or single limb mini-squat (Ageberg, et al., 2010). Common dual limb functional tests include the mini squat and partial squat or small knee bend (SKB) originally described by Sahrmann (2002). The SKB (double

126

leg) and the single leg SKB are common lower extremity functional tests (Reid, et al., 2003). The SKB test is described by Sahrmann (2002) as part of a lower quarter examination and the single leg SKB has been described as a specific screening test by Mottram and Comerford (2008). All of these double and single leg tests involve hip and knee flexion and although similar they have differences in protocols that are likely to present different movement challenges. It appears to be common clinical practise, when indicated, to further evaluate dynamic trunk and lower extremity alignment using a lunge (Crossley, Cook, Cowan, & McConnell, 2006; Thijs, Van Tiggelen, Willems, De Clercq, & Witvrouw, 2007) or hop lunge (Cook, 2006). Tests used to evaluate movement quality/dynamic alignment in a similar manner have included the lunge and hop lunge (Whatman, et al., 2011a), single leg step downs (Earl, Monteiro, & Snyder, 2007) and drop jumps/landings (Ford, et al., 2007). Many of the functional tests are used with athletes (Chiaia, et al., 2009; Reid, et al., 2003) and also in general clinical practice by physiotherapists and sports physicians to assess less active individuals. Tests such as the SKB and SLS may be more appropriate in the clinical space for injured athletes and less active individuals than higher demand tests such as hopping and drop jumps.

Reliability of lower extremity kinematics in functional screening tests

A key to the use of lower extremity screening tests clinically is the reliability of the participant's joint kinematics on repeat tests. Within-subject kinematic variation (both within and between-days) is the most important type of reliability for clinicians to consider if they intend to use the tests to monitor performance of their clients (Hopkins, 2000). It is crucial to know if the kinematics are consistent enough from test to test and day to day for making clinical decisions. For example, a clinician needs to know if following a rehabilitation programme the change in knee alignment noted visually during a lower extremity functional test is real or whether the change is due to expected kinematic variability with repeat testing. The reporting of absolute variability (such as typical error in degrees) provides greater clinical meaning than the more commonly reported relative variability (such as ICC). The reliability of these tests also needs to be established if they are to be used in longitudinal studies evaluating injury risk or the effect of rehabilitation interventions. This also necessitates investigation of the reliability over timeframes longer than a few days or a week as sporting seasons can go on for many weeks. In combined biomechanical and epidemiological studies, biomechanical variables

measured at the start of the season need to demonstrate sufficient reliability over the length of the season if they are to be considered valid risk factors for subsequent injuries recorded during the season (Ford, Myer, & Hewett, 2007). Additionally Hopkins (2000) suggested the typical error used for estimates of sample size and individual differences in experiments needs to come from a reliability study of the same duration as the experiment. Reliability across a range of ages needs consideration given that functional tests are used not only for assessment of adults but also for screening for risk of injury in younger athletes (children and adolescents). With their developing neuromuscular control, younger age groups may show greater variation in functional test performance than adults.

The reliability of joint kinematics during lower extremity functional tests reported in the literature is summarised in Table 8.1. There were only two studies investigating the reliability of kinematics during a SKB or lunge (Whatman, et al., 2011a; Whatman, Hume, & Hing, 2012a) and only one study for the hop lunge (Whatman, et al., 2011a). Other studies reporting the reliability of joint kinematics in lower extremity functional tests were limited by a lack of three dimensional (3D) motion analyses, only one report on young athletes performing tests other than the drop jump (Whatman, et al., 2012a), few studies considering the trunk and pelvis, studies only assessing the dominant side, few reports of between-day reliability (particularly over periods longer than a week) and a lack of absolute measures of reliability in degrees. A variety of test protocols have been used (see Table 8.1) with some more suitable for clinical use and better able to replicate a functional position than others. For several protocols reliability has not been reported.

There were few studies reporting the reliability of joint kinematics during the SLS, single leg SKB or single limb mini squat (all similar but not identical tests). Three dimensional motion analyses is the gold standard for assessing joint kinematics, however only one study (Zeller, et al., 2003) reported the reliability of 3D trunk and lower extremity joint (hip, knee and ankle) kinematics during the SLS. Three studies used 2D techniques to assess the reliability of frontal plane kinematics (Levinger, Gilleard, & Coleman, 2007; Stensrud, Myklebust, Kristianslund, Bahr, & Krosshaug, 2011; Willson, Ireland, & Davis, 2006). Willson et al., (2006) reported the knee frontal plane projection angle (FPPA) was reliable within-day (ICC=0.88) but did not report between-day reliability, while Levinger et al., (2007) reported acceptable reliability within and between-days (ICC 0.88 and 0.74). Levinger et al., (2007) is

the only author to have reported between-day reliability (1 week apart) of the SLS and to report an absolute measure of reliability (between-day SEM=1.7°). Both these studies limited the SLS to 45° knee flexion, while Stensrud et al., (2011) allowed knee flexion to 90° and reported similar within-day reliability for the left leg (ICC=0.84) but not the right leg (ICC = 0.57). This highlights the need for studies investigating both limbs. The 2D knee FPPA has been shown to be moderately associated with some hip and knee 3D kinematic variables but not others including 3D knee valgus (Whatman, Hume, & Hing, 2012c; Willson & Davis, 2008b). Several other studies (Ageberg, et al., 2010; DiMattia, Livengood, Uhl, Mattacola, & Malone, 2005, Crossley et al., 2011; Pantano, White, Gilchrist, & Leddy, 2005; Zeller, et al., 2003) reported the use of a variety of protocols that included differences in the amount of knee flexion, foot position, arm position, head position and movement tempo. Some studies used a small box (Crossley, et al., 2011) or allowed finger tip balance (Ageberg, et al., 2010). The different protocols may present different challenges to neuromuscular control and result in different movement patterns that influence reliability. Therefore generalisation of reliability reported from studies is not appropriate unless the protocol conditions are similar.

To standardise the performance of SLS, several studies have relied on monitoring the amount of knee flexion (Claiborne, et al., 2006; Levinger, et al., 2007; Willson, et al., 2006). Regulation of knee flexion angle is not common clinically due to the extra time and equipment required and thus the inclusion of additional monitoring in a test protocol may not give an indication of reliability that is likely achievable in the clinic. Ageberg et al., (2010) had participants looking at the position of the anterior aspect of their knee relative to tape on the floor to try to standardise the amount of knee flexion. While eye focus on a target is a simple technique aiding clinical use, and possibly reliability (although this was not reported), this alters the natural head/trunk posture and thus may not be the most relevant assessment of movement quality/dynamic alignment.

There were only two studies (Whatman, et al., 2011a; Whatman, et al., 2012a) reporting the reliability of trunk, pelvis and lower extremity 3D kinematics during a SKB and single leg SKB. In a group of healthy adults Whatman et al. (2011a) showed reliability of peak 3D kinematics was acceptable for the majority of kinematic variables of interest to physiotherapists (ICC ≥0.80). This was the only study to include the reliability of the trunk (ICC = 0.46 to 0.97) and pelvis (ICC = 0.70 to 1.0) and used

simple clinical instructions to standardise the depth of the hip and knee flexion without the need for additional equipment. In the only study of reliability of lower extremity kinematics during a SKB in young athletes these same authors showed the majority of 3D kinematics were reliable (ICC ≥0.60) over a longer period of eight to ten weeks (Whatman, et al., 2012a).

Reliable hip and knee joint 3D kinematics during a variety of similar step-down tasks has been reported for within-day (ICC >0.7, SEM = 0.5 to 2.2°) and between-day (ICC=0.75 to 0.88, SEM=1 to 4°) (Bolgla, Malone, Umberger, & Uhl, 2008; Earl, Hertel, & Denegar, 2005; Earl, et al., 2007; Whatman, et al., 2011a). As with the SLS, between-day reliability for the step-down has not been reported over periods of longer than one week. Whatman et al. (2011a) reported acceptable reliability of peak 3D trunk and pelvis kinematics during the step-down within-day (ICC = 0.79 to 1.0) and between-days (ICC = 0.46 to 0.99).

Only two studies have reported the reliability of 3D kinematics during a lunge and hop lunge (Alkjær, Henriksen, Dyhre-Poulsen, & Simonsen, 2009; Whatman, et al., 2011a). Using a lunge technique described for clinical assessment of dynamic alignment (Crossley, et al., 2006; Thijs, et al., 2007) reliability within-day and between-days for trunk, pelvis and lower extremity frontal and transverse plane kinematics appears acceptable (Whatman, et al., 2011a). Using a lunge protocol with hip and knee flexion to 90°, and keeping the contralateral leg on the ground which is more common in training (Kritz, Cronin, & Hume, 2009), between-day knee flexion has been shown to be moderately reliable (ICC = 0.53 to 0.69; ME = 4.4° to 5.8°) (Alkjær, et al., 2009).

Several studies reported the reliability of kinematics during the drop jump/landing task with three studies (Earl, et al., 2007; Ford, et al., 2007; Whatman, et al., 2012a) using 3D motion analysis. Ford et al., (2007) reported the reliability of peak 3D joint lower extremity kinematics, within-day [ICC = 0.93 to 0.99; typical error (TE) = 0.9° to 3.2°] and between-day seven weeks apart (ICC = 0.60 to 0.92; TE = 1.3° to 5.5°). Between-day reliability of peak 3D joint kinematics one week apart has also been reported as good (ICC >0.84) (Earl, et al., 2007). Most recently, Whatman et al (2012a) reported the majority of peak 3D kinematics were reliable during the drop jump within-day (ICC = 0.67 to 0.99) and between-days (8 to 10 weeks apart) (ICC = 0.27 to 0.89). In contrast to the study of Ford et al.,

(2007), Whatman et al., (2012a) reported on both legs separately. Noyes, Barber-Westin, Fleckenstein, Walsh and West (2005) used a 2D video analysis technique to report the between-day reliability of hip, knee and ankle separation when landing in a drop jump (ICC = 0.94 to 0.96). Furthermore Stensrud et al., (2011) reported within-day reliability of the 2D frontal plane knee angle when landing from a drop jump (ICC = 0.58) and single leg drop jump (ICC = 0.58 and 0.70). Unlike the studies of the other functional tests discussed earlier, studies investigating the drop jump used a range of population ages including school children, young athletes, adolescents and adults. However, again there were only two studies (Ford, et al., 2007; Whatman, et al., 2012a) that reported absolute reliability in degrees (within-day 0.4° to 3.6°, between-day 1.2° to 5.7°), which is most meaningful to clinicians.

It should be acknowledged that errors in the placement of markers on participants undergoing 2D and 3D biomechanical analyses can alter the levels of reliability of joint kinematics (Ford, et al., 2007) particularly when measurements are taken on different days. The reliability can also be affected if there is excessive marker movement during the test (McLean, et al., 2005). All of the studies reviewed used experienced clinicians to place markers and used recommended marker attachment methods minimising the effect of marker placement error. Therefore any variability in kinematics is most likely due to the participant rather than the biomechanical assessment technique although marker placement error still needs to be considered as a limitation of kinematic repeatability.

In summary, reliability of lower extremity joint kinematics during functional tests ranged from fair to excellent. However there were only a few studies that used 3D motion analysis techniques, that reported absolute reliability in degrees and that assessed between-day reliability, particularly over periods longer than one week. The reporting of trunk and pelvis kinematics was limited and the reliability of tests such as the lunge and hop lunge received limited attention. Additionally, studies used a variety of protocols, some better replicating actual functional activities (e.g. running and landing) than others and with some protocols less likely to be appropriate for clinical use. A key consideration is simple methods for standardisation of hip and knee flexion range of motion as this is likely to influence transverse and frontal plane kinematics of all segments and thus the outcome of the test (Kernozek, Torry, van Hoof, Cowley, & Tanner, 2005; Zeller, et al., 2003). Simple instructions for

clinical use have been reported for a range of functional tests in adults and young athletes (Whatman, et al., 2011a; Whatman, et al., 2012a). A further limitation of the studies to date is that the majority used healthy, active participants of a similar age. Further research is required to determine the reliability in other populations including those with pathology. Further investigation of children and adolescents is particularly important given the increasing use of these types of lower extremity functional tests in these age groups. In younger participants the complexity of the task may result in increased variability of movement pattern which could limit the use of some tests. Given most studies have looked at the dominant leg only further studies are required of both limbs including comparisons between dominant and non-dominant limbs.

Table 8.1: Reliability of kinematics during lower extremity functional screening tests.

Study	Functional test	Participants and test protocol	Main kinematic outcome measure	Reliability
Levinger et al. (2007)	Single leg squat	Females in early 20's. Single leg squat (3 trials, 1 week apart) with knee flexion of approximately 45°. The toe tips of the left leg touched the floor with the heel raised in order to support the movement and to prevent imbalance during the squat. The subject stood in a comfortable stride step with the knee to be measured forward and arms at the sides. Participants received instructions to keep their trunk upright while performing the squat. The participants performed the squat test slowly to verbal instruction.	Medial displacement of the knee as measured by the femoral frontal plane angle (2D) and femoral deviation, both at a knee flexion angle of 45°.	Within-day: ICC = 0.60 to 0.88. Between-day: ICC = 0.46 to 0.74, SEM = 1.7°.
Willson et al. (2006)	Single leg squat	Male and female athletes (mean age 19 yrs). An adjustable stool was placed behind the participants at a height that represented the distance from the floor each subject would need to assume to achieve 45° of knee flexion. Participants were asked to squat on the interested leg until they lightly touched the stool with their seat.	Frontal plane projection angle (2D) at 45° knee flexion.	Within-day: ICC = 0.88.
Stensrud et al. (2011)	Single leg squat	Female handball players (mean age 20 yrs). Single leg squat: Participants squatted to 90° knee flexion with hands on hips while focusing straight ahead (a trial was not valid if the other leg was held in front of the body during the squat, if it touched the ground, if the hands were removed from the side of the body, or the player looked down or fell).	Frontal plane knee angle (2D) at maximum knee flexion.	Within-day: (R) ICC = 0.57, (L) ICC = 0.84.
Crossley et al. (2011)	Single leg squat	Healthy adults (mean age 24 yrs). Participants stood on a 20 cm box, arms folded across chest, instructed to squat down as far as possible (rate of 1 squat per 2 s). Pictures show contralateral leg held out in front of box in a manner similar to a step-down.	Not measured.	Not reported.
Zeller et al. (2003)	Single leg squat	Male and female athletes (mean age 20 yrs) stood on their dominant leg with their arms crossed over their chest. They squatted down as far as possible and then returned to single leg stance without losing their balance, within 5secs.	Maximum range of 3-dimensional trunk and lower limb (hip, knee, ankle) angles.	Not reported.
Pantano et al. (2005)	Single leg squat	Males and females (18-29 yrs).Right foot was placed on a step pointing straight ahead while they stepped down onto the lower step with their left foot. Knee flexion was limited to 45°.	Peak knee valgus angle (2D).	Not reported.
DiMattia et al. (2005)	Single leg squat	Healthy adults (mean age 24 yrs) stood on their stance leg, while they lifted their non stance leg off the ground so their hip flexed to approximately 45° and their knee flexed to approximately 90°. Their shoulders were flexed to 90° and elbows extended with hands clasped together in front. They squatted down to 60° knee flexion, and returned to the starting position in less than 6 s.	Peak knee valgus/varus, Hip ab/adduction angles (2D).	Not reported.
Weir et al. (2010)	Unilateral squat	Males (age >18 yrs) stood on right leg, the trunk was upright, without rotation or lateral flexion, and the contralateral leg was positioned with the hip in neutral position and the knee in 90° flexion. Participants moved at a self-selected pace into a squat position.	Not measured.	Not reported.
Chmielewski et al. (2007)	Unilateral squat	Males and females (mean age 22 yrs) stood on one leg, trunk upright and contralateral leg with hip neutral and knee flexed 90°. Participants squatted to 60° knee flexion at a self selected pace.	Not measured.	Not reported.
Ageberg et al. (2010)	Single limb mini squat	Females (age 18-37 yrs), "T" marked on floor, foot aligned with stem of "T" and 2nd toe on stem, finger tip support from bar, instructed to look down and bend knee, without bending forward from hips, until could no longer see the line across the toes (approximately 50° knee flexion). Speed of 20 squats/min rate set by metronome (1 squat 3 s). Contralateral hip slight flexion, knee approximately 80° flexion.	2D knee valgus, 3D hip and knee joint angles at peak knee flexion.	Not reported.
Örtqvist et al. (2011)	Single limb mini squat	Children (age 9-16 yrs). Protocol as for the study by Ageberg (2010), except subjects had to perform as many knee flexions as possible in 30 s.	Not measured.	Not reported.

Sahrmann (2002)	SKB-partial squat	Bilateral hip and knee flexion to 45° with heels on floor. Path of the knee should stay over longitudinal axis of the foot (2nd toe).	Not measured.	Not reported.
Reid et al. (2003)	Single leg SKB	Stand on one leg, slowly bend knee keeping patella over 2nd and 3rd toes, heel remains on ground.	Not measured	Not reported.
Reid et al. (2003)	Hop land	Standing on one leg, jump approximately 1m to land on a flexed knee.	Not measured	Not reported.
Bolgla et al. (2008)	Step-down	Females (mean age 25 yrs). Stair stepping task (5 practice trials, standardized rate of 96 beats per minute, 1 week apart). Participants were instructed to ascend and descend two 20-cm high steps with a minimum of 3 strides prior to and immediately following stair stepping.	Average 3D, hip and knee frontal and transverse plane angles.	Between-day: ICC = 0.75 to 0.88. SEM = 1° to 4°
Chmielewski et al. (2007)	Lateral step-down	Males and females (mean age 22 yrs). Stood on one leg on the edge of an adjustable step, trunk upright and contralateral leg unsupported. Participants lowered themselves at a self selected pace until the contralateral heel contacted the ground.	Not reported.	Not reported
Earl et al. (2005)	Lateral step-down	Males and females (mean age 21 yrs). Lateral step-down off a 20.3-cm block (5 trials, standardized rate of 60 beats per minute). Participants were instructed to "take a normal step forward with your test leg, then, keeping your pelvis level and your thigh straight lower the opposite foot to lightly tap the floor and return to a standing position."	Peak 3D eversion, tibial internal rotation, knee valgus (abduction), knee flexion, hip adduction, and hip internal rotation.	Within-day: ICC >0.7 SEM = 0.5° to 2.2°.
Earl et al. (2007)	Step-down	Males and females (mean age 22 yrs). Step-down (3 trials, 1 week apart) - participants were instructed to "Stand on your right leg and slowly lower the left foot to lightly touch the floor with your heel, and return to standing on both feet."	Peak 3D joint angles of rearfoot eversion, knee flexion, knee internal rotation, knee abduction, hip internal rotation, and hip adduction.	Between-day: ICC >0.87
Alkjær et al. (2009)	Lunge	Males and females (mean age 26 yrs). Forward lunge (1 to 4 trials, 1 week apart) onto a force plate, flexing the knee to 90° and subsequently extending the knee in order to return to the starting position. Upper body perpendicular to the ground and contralateral foot in contact with the ground during the entire movement.	Peak knee flexion angle (3D).	Between-day: ICC = 0.53 to 0.69. ME = 4.4° to 5.8°
Ford et al. (2007)	Drop jump	Middle and high school, male and female soccer players. Participants performed a drop jump (3 trials, 7 weeks apart) starting on top of a 31-cm box with their feet positioned 35 cm apart. They were instructed to drop directly down off the box and immediately perform a maximum vertical jump.	Peak landing hip, knee and ankle joint angles (3D).	Within-day: ICC = 0.93 to 0.99; TE = 0.9° to 3.2°. Between-day: ICC = 0.60 to 0.92; TE = 1.3° to 5.5°.
Noyes et al. (2005)	Drop jump	Male and female adolescent athletes. Participants performed a jump land sequence (3 trials, 1 week apart). They jumped off a box, landed, and then immediately performed a maximal vertical jump.	2Dvideo analysis. Hip, knee and ankle separation distance on landing.	Between-day: ICC = 0.94 to 0.96.
Earl et al. (2007)	Drop jump	Males and females (mean age 22 yrs, 3 trials, 1 week apart). Participants were instructed to "Drop off the front of the block, land with both feet on the force plate, and then immediately jump as high as possible." Participants aligned their toes with tape marks placed 35 cm apart to standardize the take-off position, but landed with the feet in a natural alignment.	Peak 3D joint angles of rearfoot eversion, knee flexion, knee internal rotation, knee abduction, hip internal rotation, and hip adduction.	Between-day: ICC >0.84.
Stensrud et al. (2011)	Drop jump, Single leg drop jump	Female handball players (mean age 20 yrs). Single leg drop jump: Participants dropped off a 10 cm box with one leg before immediately performing a maximum vertical jump on the same leg, moving their arms freely. Drop jump: Performed as described by Hewett (2005).	Frontal plane knee angle (2D) at maximum knee flexion.	Within-day: Single leg drop jump (R) ICC = 0.58, (L) ICC = 0.70; Drop jump ICC = 0.89.
Kernozek et al. (2005)	Drop landing	Male and female athletes (mean age 24 yrs). The landing task consisted of dropping 60 cm from an adjustable "hang bar" suspended over the force plate.	Hip, knee and ankle joint angles (3D). Angle at ground contact	Within-day: ICC = 0.79 to 0.93 (based on pilot work).

			and difference between angle at ground contact and maximum knee flexion).	
Whatman et al. (2011a)	SKB, Single leg SKB, Lunge, Hop Lunge, Step-down	Males and females (mean age 22 yrs), 3 trials, 1 to 2 days apart. SKB - trunk maintained in an upright position - continued SKB until reached maximum dorsiflexion without lifting their heels, Lunge - lunge forward a distance of approximately one and a half times the length of their normal gait stride (continue the lunge until reaching maximum dorsiflexion of the stance leg without lifting their heel), Hop lunge - jump forward a distance of approximately 1.0 m and on landing flex the hip and knee, Step-down- stepped slowly down from a step 20 cm high.	Peak 3D kinematics – trunk and pelvis lateral flexion and rotation, hip flexion, adduction, internal rotation, knee flexion, abduction, ankle dorsiflexion, eversion	Within-day: ICC ≥0.92, TE ≤1.3°. Between-day: ICC ≥0.80. TE ≤3.9°.
Whatman et al. (2012a)	SKB, Single leg SKB, Drop Jump	Young athletes (mean age 11yrs), 3 trials, 8 to 10 weeks apart. SKB as above (single leg SKB right and left), Drop Jump - dropped directly off a 25 cm box and immediately jumped vertically as high as possible.	Peak 3D kinematics – hip flexion, adduction, internal rotation, knee flexion, abduction, ankle dorsiflexion, eversion	Within-day: ICC ≥0.85, TE = 0.4° to 3.6°. Between-day: ICC = 0.60 to 0.92, TE = 1.2° to 5.7°.

ICC = Intraclass Correlation Coefficient, TE = Typical Error, SEM = Standard Error of Measurement, SKB = Small Knee Bend, SLS = Single Leg Squat, PFPS = Patellofemoral Pain Syndrome.

Reliability and validity of visual assessment of functional screening tests

Frohm, Heijne, Kowalski, Svensson and Myklebust (2011) encouraged clinicians to identify dysfunctional movement patterns before pain or functional problems developed and emphasised the need for reliability and validity of the tests used. Visual assessment of movement quality/dynamic alignment during functional tests is a common part of screening athletes prior to participation in sport (Reid, et al., 2003). Screening movement quality is promoted as a way to decrease an athlete's risk of injury and/or enhance performance. Screening for abnormal movement patterns has been included as a strategy in a recent position statement on the prevention of paediatric overuse injuries in sport (McLeod, et al., 2011), however the authors did note that functional tests to assess movement patterns have yet to be validated in this population.

In a review of qualitative visual analysis of human movement Knudson (2000) discussed several factors that influence reliability and validity. Validity was affected by the velocity of the movement, difficulties with estimating joint angles (a 5° error for static posture and a 15° to 30° error for fast movements) and the choice of critical features that were assessed. The perceptual style of the observer may also be important with some raters achieving more accurate assessment by first gaining an overall impression of the critical features prior to analysing discrete body segments and raters preferring different types of scales (e.g. visual analogue versus ordinal scale) (Knudson, 1999; Morrison, 2000). Further suggestions to improve reliability include well defined critical features and simple rating methods (three point scales sufficient for rating most critical features) that remained meaningful (Knudson, 2000). Rating environment may also be important with recommendations suggesting a neutral background with horizontal or vertical references (Morrison, 2000).

Studies reporting the reliability and/or validity of visual ratings of lower extremity functional tests are summarised in Table 8.2. The majority of studies evaluate the level of agreement between raters (reliability) by reporting the Kappa Coefficient or Agreement Coefficient One (AC1) and/or percent agreement. The interpretation of Kappa/AC1 is most often based on the scale proposed by Landis and Koch (1977) where a value of 0.81 to 1.00 represents excellent/almost perfect agreement, 0.61 to 0.80 substantial/good agreement, 0.41 to 0.60 moderate agreement, 0.21 to 0.40 fair agreement, 0.01

to 0.20 slight agreement and <0.00 poor agreement. Although visual assessment is common, the reliability and validity of these ratings across many functional tests, remains variable. Reported intra-rater agreement generally ranges from moderate to excellent and inter-rater agreement from slight to good (Ageberg, et al., 2010; Chmielewski, et al., 2007; Crossley, et al., 2011; DiMattia, et al., 2005; Ekegren, Miller, Celebrini, Eng, & Macintyre, 2009; Örtqvist, et al., 2011; Whatman, Hing, & Hume, 2011b; Whatman, Hume, & Hing, 2012b). Variation in rater agreement is likely due to differences in the functional tests themselves as well as the population rated, the rating methods, variations in the amount of rater training, raters with varied experience and differences in analysis. A few studies have reported groups with good versus poor dynamic alignment determined by visual ratings show differences in 3D hip but not knee kinematics (Ageberg, et al., 2010; Whatman, et al., 2012b). There is evidence that visual ratings are most accurate at determining differences in 2D kinematics (pelvis and knee) (Ageberg, et al., 2010; Stensrud, et al., 2011; Whatman, et al., 2012b). Factors contributing to improved reliability and validity of visual assessments may include dichotomous rating scales, clearly described critical features and/or anatomical references, slower velocity tests, extensive rater training in the specific assessment technique and more clinical experience. Studies with larger numbers of raters (>10) show a greater range of intra-rater agreement (slight to excellent) but similar inter-rater agreement. The generalisability of findings in studies with small rater numbers (two or three) to the broad range of clinicians is questionable. Further development of visual assessment is thus required with consideration given to the test protocol, the rating method, the critical features assessed, the velocity of the test being rated and adequate groups of raters.

Inter-rater agreement

Visual rating of the SLS and single limb mini squat has received the most attention in the literature. There is however little consistency in the visual rating method used or the method of evaluating the reliability and validity of the ratings. The best inter-rater reliability ($\kappa = 0.92$, 96% agreement) has been reported by Ageberg (2010) in a study that provided explicit guidelines and training for the two experienced physiotherapists who used a simple dichotomous rating scale of one body segment and anatomical references (knee relative to foot). All these factors and the experienced nature of the small number of physiotherapists likely helped to achieve the high level of agreement. Experience did not

affect the lower inter-rater agreement (AC1 = 0.22 to 0.71) in two recent studies where no specific training was provided for rating of movement quality/dynamic alignment and larger numbers of physiotherapists were used (44 and 66) (Whatman, et al., 2011b; C. Whatman, et al., 2012b). Studies with larger numbers of physiotherapists are more generalisable to a range of physiotherapists in clinical practice. Additionally rater training could have raised agreement above that which could be expected in clinical practice where clinicians usually do not have specific education on visual rating. A similar dichotomous rating method used to assess young participants (9 to 16 yrs) performing multiple repetitions of a single limb mini squat in 30 s showed lower agreement between a physiotherapist and orthopaedic surgeon ($\kappa = 0.57$, 79% agreement) in the study by Örtqvist (2011). Training was not given prior to this study which may give a better indication of actual agreement likely in current clinical practice although the different background of the raters may have also had an influence. Given the requirement to perform as many single limb mini squats as possible in 30 s the velocity of the movement was also likely greater. Faster movements have been reported to be more difficult to rate than slower movements (Knudson, 2000). Velocity of movement has been shown to influence the ability to rate knee position in young athletes during a drop jump (Whatman, et al., 2012b). The dichotomous rating of knee position relative to the foot was also used in the only study that has reported inter-rater agreement when rating young athletes in a single leg SKB (AC1 = 0.52) (Whatman, et al., 2012b).

Ekegren (2009) also reported good inter-rater agreement ($\kappa = 0.77$ to 0.80) using a simple dichotomous method to rate knee position in adolescent female soccer players during a drop jump. A comparable rating method was used by Whatman et al. (2012b) to rate younger athletes, however inter-rater agreement was not as good (AC1 = 0.34) and this may again have been partly due to the limited training/instruction given to the raters and/or to the larger group of 66 raters used in this study. Dichotomous rating of knee position in the frontal plane was common in several studies however the described reference points used appear variable. Some studies have been more specific rating the patella relative to the first toe (Ekegren, et al., 2009), or the knee relative to the second toe (Ageberg, et al., 2010), or the knee relative to the first and second ray (Tofte, Tillman, & Chmielewski, 2011). Others have been more general, rating the knee relative to the middle of the foot (Crossley, et al., 2011) or ankle (Örtqvist, et al., 2011). The rating of knee position relative to the second toe during the

single limb mini squat achieved the highest inter-rater agreement (Ageberg, et al., 2010) but again this was for only two raters. Moderate inter-rater agreement (AC1: 0.41 to 0.54) has also been reported for dichotomous ratings of the pelvis during single leg SKB, lunge and hop lunge (Whatman, et al., 2011b; Whatman, et al., 2012b).

Visual rating in the clinic commonly involves assessment of more than one body segment. Mottram (2008) recommended rating the trunk, pelvis, knee and foot when assessing lower extremity movement quality. Overall ratings classifying movement deviations as minor, moderate or marked are also common in clinical practice (Reid, et al., 2003) and likely more difficult than rating a single segment. Chmielewski (2007) used a method that required rating multiple segments based on deviations from a neutral position and segment oscillations, more in keeping with current clinical practice but this resulted in low inter-rater agreement (κ = 0.23 to 0.53, 20-50% agreement). Similar agreement using a segmental method (4 point ordinal scale) was reported by Whatman et al. (2011b) across a range of tests including SKB, Single leg SKB, lunge and hop lunge (AC1 = 0.25 to 0.42) for a group of 44 physiotherapists. Segmental oscillations and deviations from neutral during a SLS were also rated on a four point scale by Weir (2005) who reported low agreement for a small group of physiotherapists and sports physicians. Poor agreement was also reported by DiMattia (2005) who required ratings on a four point scale based on reference hip and knee angles. Visual estimation of angles is known to be difficult (Krosshaug, et al., 2007) and to not help visual rating of movement (Knudson, 1999). Estimating angles does not appear to be common clinically in this type of lower extremity assessment.

In spite of this lower inter-rater agreement, reported with more complex rating methods, Piva (2006) managed to achieve good inter-rater agreement (κ = 0.67, 80% agreement) using a segmental (arms, trunk, pelvis and knee) rating approach. This study rated movement in the lateral step-down which involves a pattern very similar to the single limb mini squat. However, the study included only four raters (of varied experience) who all received training and a detailed manual explaining the rating procedure which likely aided agreement. In addition, unlike the other studies reviewed this study reported visual agreement when rating injured participants with patellofemoral pain syndrome. The only other study reporting inter-rater agreement focused solely on the quality of knee flexion giving the

results limited application in screening for many overuse injuries (von Porat, Holmstrom, & Roos, 2008). However this was the only study to record visual observations on a horizontal 11-point rating scale. Although this method of rating needs further investigation, the method produced moderate inter-rater agreement (ICC = 0.57 to 0.76) and presents an alternate rating scale that some clinicians may find useful.

Intra-rater agreement

Intra-rater agreement, when visually rating the SLS and single limb mini squat, has been reported in three studies over periods of approximately one week using two or three raters. The highest agreement (κ = 0.61 to 0.80, 73-87% agreement) was reported recently by Crossley et al., (2011) for experienced physiotherapists in a study where example ratings were provided as training prior to repeat ratings made one week apart. Providing examples and rater training likely contributed to the substantial agreement reported. As the authors referred to digital images it was unclear whether the ratings were made from still images or video and how many repetitions of video were viewed. Nevertheless the rating method used involved a relatively complex evaluation of the trunk, pelvis and knee in a manner similar to common clinical practice. This study also showed that two physiotherapists with musculoskeletal postgraduate qualifications and more experience achieved higher agreement than a graduate physiotherapist. This influence of experience on intra-rater agreement was also reported by Whatman et al. (2011b). Chmielewski (2007) used a similar segmental rating method (with less detailed criteria and including a rating of segment oscillation) and reported lower agreement (κ = 0.35 to 0.68, 32% to 60% agreement) for ratings separated by an average of 10 weeks. This study also used an overall rating method (ratings based on a general impression of movement quality across all segments) which resulted in similar intra-rater agreement (κ = 0.13 to 0.50, 56% to 76% agreement). Using a similar segmental method to rate a range of movements (SKB, single leg SKB, lunge, hop lunge) over three to four weeks, 33 physiotherapists showed a wide range of intra-rater agreement (AC1 = 0.01 to 0.96), (Whatman, et al., 2011b). Intra-rater agreement using a simple dichotomous rating and anatomical reference points (knee relative to ankle) to rate SKB, single leg SKB and drop jump in young athletes and healthy children has generally shown better agreement (κ = 0.48; AC1 = 0.56 to 0.81) (Örtqvist, et al., 2011; Whatman, et

al., 2012b). Agreement in these studies when rating children and young athletes was similar to that achieved with adults.

Intra-rater agreement for rating the unilateral squat and lateral step-down over a five week period has been reported as poor (ICC = 0.55, 49% agreement) for a group of experienced sports physicians and physiotherapists (Weir, 2005). In this study the segments rated and the criteria for rating were not well defined. Intra-rater agreement has also been reported for visual observation of the drop jump (Ekegren, et al., 2009; Whatman, et al., 2012b). Both studies used a simple dichotomous rating (knee position relative to the foot). Ekegren (2009) reported good to excellent agreement (κ = 0.75 to 0.85, 86-93% agreement), the experienced nature of the physiotherapists and training again likely aided agreement in their study. Whatman et al. (2012b) reported similar agreement but with more variation in a group of 26 physiotherapists (without additional training) rating young athletes (AC1 = 0.14 to 0.92).

The reliability of visual rating of lower extremity movement quality/dynamic alignment utilising the use of video pause and rewind has also been reported, however this method does not seem common in current clinical practice, as it is more time consuming and does not afford the benefit of immediate evaluation of changes in performance. Padua (2009) and Onate (2010) investigated the reliability and validity of the Landing Error Scoring System (LESS), another screening tool for evaluating lower extremity movement patterns in a jump landing task. The LESS was designed for identifying individuals at increased risk for ACL injury. It is a score based on a count of 17 landing technique errors (assessing trunk and lower extremity using dichotomous ratings), visually rated from both frontal and sagittal plane video. Raters who scored the LESS underwent an extensive training programme. Padua (2009) and Onate (2010) both reported good inter-rater agreement (ICC = 0.84) for the overall LESS, while Padua (2009) also reported high intra-rater agreement (ICC = 0.91). Additionally Onate (2010) reported intra-rater reliability for individual LESS criteria (κ = 0.46 to 1.0, 65% to 100% agreement). Interestingly the two criteria that achieved perfect agreement were both ratings of the knee position relative to the foot. While direct comparison is inappropriate due to methodological differences, the use of slow motion and pause may offer better reliability and/or validity for some but not all types of visual ratings.

Validity of visual assessment

There is evidence supporting the validity of visual ratings of lower extremity movement quality/dynamic alignment during functional screening tests (see Table 8.2), however the methods are variable and limited. Assessment of validity has been based on evaluation of kinematic differences between visually rated groups, comparison to consensus/expert visual ratings and the ability to estimate discrete joint angles. Several studies have investigated the validity of visual ratings of the SLS or single limb mini squat. In the only study comparing visual ratings to 3D and 2D motion analysis, Ageberg (2010) reported that participants visually rated as knee medial to foot had significantly greater hip internal rotation but not knee valgus (as measured with 3D motion analysis) and greater medial knee position (as measured by 2D motion analysis). The authors also claimed the visual assessment could discriminate between those with and without a medial knee position based on the area under the receiver operating characteristic (ROC) curve (area under the curve for knee valgus in 2D = 0.87). Given that only two raters were involved and there was no use of video slow motion or replay there may be some question as to the accuracy of the visual ratings that were used as the criterion measure. Inter-rater agreement was however high (κ = 0.92). Tofte et al., (2011) undertook a similar study (using 2D motion analysis only) where one clinician grouped participants (based on knee relative to foot position) performing a lateral step-down. These groups were reported to have significantly different 2D knee FPPA. Both studies showed the ability of visual ratings to differentiate groups based on kinematics (stronger association between visual and 2D than 3D) that may place them at greater risk of injury. However, both involved only one or two raters and extensive rater training including example ratings prior to the study. This level of validity may not be what is currently achievable clinically across a broader range of clinicians. A further study where one physiotherapist rated 186 female handball players also reported that groups determined by visual ratings showed significantly different 2D knee FPPA as measured from video (Stensrud, et al., 2011).

Recently, Whatman et al. (2012b) concluded the dichotomous visual ratings of the experts (using video pause and replay, but without additional training or example ratings) for SKB, single leg SKB and drop jumps were able to differentiate young athletes with different hip and pelvis kinematics (measured using 2D and 3D motion analysis techniques) that might place them at increased risk of injury. Again visual ratings differentiated 2D kinematics much better than 3D kinematics. Ekegren,

Miller, Celebrini, Eng and Macintyre (2009) also looked at the validity of visual ratings compared to 3D motion analysis during a drop jump. Knee valgus angles were compared to their dichotomous visual ratings (knee position relative to first toe) resulting in sensitivities of 67% to 87% and specificities of 60% to 72%. Sensitivity was considered inadequate and would result in too many at-risk individuals failing to be detected. This lack of sensitivity was again in spite of the significant rater training given before the study commenced. Again only one physiotherapist was used to provide the criterion visual rating however in this case video slow motion, pause and rewind as well as multiple rating occasions were used to come to a final decision.

In an alternate approach Crossley et al., (2011) used a panel of five expert physiotherapists to gain consensus on visual ratings of the SLS. Trunk, pelvis, knee and overall movement were rated on a three point scale based on detailed qualitative criteria in a manner in keeping with clinical practice. Three physiotherapists made their own independent ratings and validity was based on good concurrency reported with the consensus panel (k = 0.60 to 0.80, 73% to 87% agreement). Interpretation of validity based on a single agreement coefficient fails to differentiate the accuracy of positive versus negative ratings. This is important information when deciding whether or not to use a rating method and/or make improvements. This is the only study however that has investigated validity with such a complex rating method. Importantly the study results also showed that people rated as having poor performance on the SLS had worse hip muscle function (delayed EMG onset in a step-up task). Whatman et al. (2012b) also compared the dichotomous knee and pelvis ratings (during the SKB, single leg SKB and drop jump) of 63 physiotherapists to those of an expert panel. The results of this study may be more generalisable to general physiotherapy clinical practice and the study addressed the accuracy of both positive and negative ratings by reporting acceptable sensitivity (≥80%) and specificity (≥50%). Clinical experience (diagnostic odds ratio 1.6 to 2.8 times better) and slower movement velocity (diagnostic odds ratio 4.9 times better) were possibly factors in better rating accuracy.

In yet another approach an earlier study by DiMattia, Livengood, Uhl, Mattacola and Malone (2005) assessed the ability of two athletic trainers to evaluate movement during a SLS based on their ability to pick discrete hip and knee reference angles. Accuracy was determined via low to moderate

sensitivity (0.22 to 0.54) and moderate to high specificity (0.58 to 0.87). The low sensitivity in this study means participants with poor movement were likely to get rated as acceptable and not receive needed intervention. The poor validity demonstrated was not surprising given the difficulties with visual assessment of joint angles already discussed.

Onate, Cortes, Welch and van Lunen (2010) and Padua et al., (2009) evaluated the validity of the LESS in which raters were aided by the use of video pause and rewind. Padua et al., (2009) concluded that participants with high LESS scores (poor landing technique) had significantly different kinematics (measured by 3D motion analysis) compared to those with low LESS scores (good landing technique). Again it appears only one rater provided the LESS scores rather than a consensus panel which would have been a stronger methodology. Onate et al., (2010) assessed validity of each component of the LESS relative to 3D motion analysis and found that validity was item dependent. Of note the knee valgus rating, a key variable in many of the other functional tests, showed only moderate agreement (74%) with 3D motion analysis.

A possible issue for the validity of visual ratings is the uncertainty in the link between 2D kinematic and 3D kinematic measures. Visual ratings often attempt to detect individuals with 3D lower extremity valgus, a recognised at risk injury position. Evidence suggests however visual ratings only clearly differentiate 2D kinematics. A concern is the relationship between 3D kinematics and 2D measures of frontal plane knee angle (McLean, et al., 2005). The most commonly reported 2D measure of knee angle is the knee FPPA measured from standard digital video recorded from an anterior view (Herrington, 2009; Olson, Chebny, Willson, Kernozek, & Straker, 2011; Willson & Davis, 2008b). This is the angle measured between a line connecting the ASIS and the mid-point of the femoral condyles and a second line connecting the mid-point of the femoral condyles to the mid-point of the malleoli. Associations between 2D knee FPPA and 3D kinematics (hip and knee angles) during SLS, SKB and drop jump have been reported as small to moderate (r = 0.21 to 0.48) (Whatman, et al., 2012c; Willson & Davis, 2008b). Olson et al., (2011) concluded that exercises to improve lower extremity alignment may improve knee FPPA but this may not be related to specific changes in 3D joint kinematics. Ekegren et al., (2009) reported that 3D joint angles in a drop jump were consistently smaller than corresponding angles measured from video due to the knee flexion and hip internal

rotation that produced knee valgus being more obvious on video in the frontal plane than in the orientation of the knee joint axis. Further studies are required to investigate the relationship between 2D measures of lower extremity alignment and 3D kinematics across a range of populations.

It is clear that the reliability and validity of visual ratings varies with the movement being rated, the method of rating and the number of raters. While agreement in several studies seems acceptable for clinical use, rater numbers were small, some studies did not report all aspects of agreement and/or there were insufficient data on some functional tests to make a judgement. Further studies using simple rating methods likely to achieve high agreement, while still providing adequate information for clinical use, are required. Based on findings to date it is likely that simple dichotomous rating methods and clearly defined anatomical references are required to achieve high agreement among raters. The movement variables that are key to evaluation of lower extremity movement quality/dynamic alignment also need further consideration as there has been a lack of consistency in previous studies. Most studies investigating inter-rater and intra-rater agreement have been limited to a small number of clinicians of similar experience and thus the generalisability of these results to all clinicians is questionable. Additionally the few studies that have attempted to quantify the effect of experience on the level of agreement have reported the effect is unclear. Although challenging, the effect of experience needs to be expressed in a manner that allows us to discuss its clinical relevance. Similar intra-rater agreement appears achievable over timeframes of more than one week versus less than one week. Agreement over one week is probably useful for screening applications but longer time periods in keeping with rehabilitation timeframes are also important. Ratings of various age groups appear similar when similar functional tests and rating approaches are used. There is some evidence that visual ratings can differentiate groups with kinematics (predominantly 2D) that may place them at increased risk of injury and that ratings agree sufficiently with those of an expert panel. The generalisability of results from some of these studies is again limited by the small number of clinicians involved. Further studies comparing ratings to 3D motion analysis and consensus expert ratings, with appropriate analysis of rating accuracy, are required to confirm this across the range of lower extremity functional tests in clinical use.

Table 8.2: Studies investigating the reliability and validity of visual rating of lower extremity movement quality/dynamic alignment during functional screening tests.

Study	Functional Test	Participants	Observational rating method	Reliability	Validity
Ageberg et al. (2010)	Single limb mini squat	Two physiotherapists rated 25 healthy subjects (18-37yrs) on 1 occasion.	Knee position relative to the 2nd toe was rated (knee over foot or knee medial to foot) over 5 reps in real time. Training in rating was given.	Inter-rater agreement κ = 0.92, PA = 96%.	Those rated as medial to foot had greater 3D internal hip rotation (p = 0.049) and greater 2-D peak tibial and thigh angle with respect to the horizontal (p=0.001). Area under the ROC curve for 2D knee valgus = 0.87.
Örtqvist et al. (2011)	Single limb mini squat	Two physiotherapists rated 33 healthy children (9-16 yrs) on 2 occasions 7 to 10 days apart.	Knee position relative to the ankle joint was rated (medial or in line) while the subject performed as many squats as possible in 30s. No training was given and ratings were done in real time.	Intra-rater agreement κ = 0.48, PA = 76% agreement. Inter-rater agreement κ = 0.57, PA = 79%	Not reported.
Crossley et al. (2011)	Single limb mini squat	Three physiotherapists rated 34 healthy adults (mean age 24 yrs) on 2 occasions, 1 week apart.	DVD of digital images rated as good, fair or poor based on qualitative criteria evaluating trunk, pelvis, knee and overall movement. Example ratings were provided. No details given on viewing of DVD's.	Intra-rater agreement κ = 0.61 to 0.80, PA 73 to 87%. No inter-rater agreement reported.	Concurrency with consensus panel (5 expert physiotherapists) 73 to 87% agreement, κ = 0.60 to 0.80. Those rated as poor showed worse hip abductor muscle function.
Weir et al. (2010)	Unilateral squat	Four sports physicians and 2 physiotherapists rated 40 males (mean age 24 yrs) on 2 occasions 5 weeks apart.	Video recordings (6 reps) were rated as poor, moderate, good, excellent based on an overall impression of deviations from a neutral alignment and segment oscillation. Rating instructions were provided.	Intra-rater agreement ICC = 0.55, 49% PA. Inter-rater agreement ICC = 0.41.	Not reported
Chmielewski et al. (2007)	SLS	Three physiotherapists rated 25 uninjured subjects (18-25yrs) on 2 occasions, 10 weeks apart.	Video of movement (3 reps). Overall and segment (trunk, pelvis, hip) rating, based on 4 point scale indicating severity of segment deviation from a neutral position and segment oscillation. Overall movement rated as good, fair, poor. Training was given, one rewind was allowed and ratings had to be completed within 30 s of movement completion.	Inter-rater agreement weighted κ = 0.0 to 0.55, PA = 20 to 82%, Intra-rater weighted κ = 0.13 to 0.68, PA = 32 to 76%.	Not reported.
DiMattia et al. (2005)	SLS	Two athletic trainers rated 50 healthy adults (26 male, mean age 24yrs) on 1 occasion.	Rated as poor, fair, good or excellent based on hip joint (flexion >65°, adduction >10°) and knee joint (valgus>10°) reference angles and balance. Ratings made in real time	Inter-rater agreement κ = 0.16 to 0.28, PA = 67 to 71%.	Compared to reference joint angles measured using 2D motion analysis - sensitivity 0.22 to 0.54, specificity 0.58 to 0.87.
Ekegren et al. (2009)	Drop jump	Three physiotherapists rated 40 healthy adolescent female soccer players on 2	DVD of landing (3 reps) was rated high or low risk based on position of patella relative to 1st toe. Given 15 s to complete ratings, no pause or rewind allowed and training was given prior to ratings.	Intra-rater agreement κ = 0.75 to 0.85, PA 86 to 93%. Inter-rater agreement κ = 0.77 and 0.80.	Compared to knee valgus angle measured with 3D motion analysis which was used as a criteria for truly

					high and low risk - sensitivity 67 to 87%, Specificity 60 to 72%.
Padua et al. (2009)	Jump landing	2 raters rated 50 military recruits (25 females) on 2 occasions, 1 week apart.	Video assistance (pause and rewind) was used to record the LESS score (a count of landing technique errors based on visual observations on a dichotomous scale), 3 jump land trials There are 17 scored items in the LESS.	Inter-rater agreement ICC = 0.84, SEM = 0.71. Intra-rater agreement ICC = 0.91, SEM = 0.42.	Subjects with high LESS scores (poor jump-landing technique) displayed significantly different 3D lower extremity kinematics compared with subjects with low LESS scores (excellent jump-landing technique).
Onate et al. (2010)	Drop jump	Two Athletic trainers.	Video assistance (pause and rewind) was used to record the LESS score (a count of landing technique errors based on visual observations on a dichotomous scale), 3 jump land trials There are 17 scored items in the LESS.	Inter-rater agreement for each item on the LESS range κ = 0.46 to 0.88. Inter-rater agreement for overall LESS ICC = 0.84.	Comparison between 3D kinematics (converted to dichotomous ratings) and LESS score, PA 10 to 100%.
Weir et al. (2010)	Lateral step-down	Four sports physicians and 2 physiotherapists rated 40 males (mean age 24 yrs) on 2 occasions 5 weeks apart.	Video recordings (6 reps) were rated as poor, moderate, good, excellent based on an overall impression of deviations from a neutral alignment and segment oscillation. Rating instructions were provided.	Intra-rater agreement ICC = 0.49, PA = 47%. Inter-rater agreement ICC = 0.39.	Not reported.
Piva et al. (2006)	Lateral step-down	Four Physiotherapists rated 30 subjects (17 female) with PFPS on 1 occasion.	Movement was rated as good, medium, poor based on a score given for quality of balance, arm, trunk, pelvis and knee movement. Ratings were made in real time and training was given.	Inter-rater agreement κ=0.67, PA 80%.	Not reported.
Tofte et al. (2011)	Heel tap (lateral step-down)	One clinician rated 44 healthy female collegiate athletes on 1 occasion.	Video of movements was rated based on knee position (0 = knee crosses midline of body, 1 = knee position is medial to first ray, 2 = knee is aligned over the first ray, and 3 = knee is aligned over or lateral to the second ray).	No reliability reported.	Groups determined by visual ratings showed significantly different knee FPPA as measured from the same video (P<0.001).
von Porat et al. (2008)	1. Knee bending 2. Step up/down 3. Crossover hop 4. One leg hop	Four physiotherapists rated 12 ACL injured subjects (mean age 40yrs) on 2 occasions.	Video of movements (combined assessment of all tests) was rated based on the quality of knee flexion (low quality=decreased knee flexion and load moved from the joint). 11 point scale used to make rating, 0 = no knee flexion, 10 = normal knee flexion. Assessment was made at normal speed and training was provided.	Inter-rater agreement ICC = 0.57 to 0.76. No intra-rater agreement reported.	Correlation to 3D knee flexion angles, Spearman's correlation coefficient=0.37 to 0.61.
Stensrud et al. (2011)	SLS, Single leg drop jump, Drop jump	One physiotherapist rate 186 female handball players (mean age 22 yrs).	Movement graded on a scale 0 to 2 based on lateral tilt of the pelvis, valgus motion of the knee and medial/lateral movement of the knee. Ratings made in real time.	NA.	Groups determined by visual ratings showed significantly different knee FPPA as measured from video (P<0.001).
Whatman et al. (2011b)	SKB, Single leg SKB, Lunge, Hop lunge	Fourty four physiotherapists rated 6 healthy adults (mean age 22 yrs).	Video of movements. Overall and segment (trunk, pelvis, hip, knee, foot) rating, based on 4 point scale. No rating instructions, ratings in real time.	Intra-rater agreement PA: 29% to 96%, AC1: 0.01 to 0.96. Inter-rater PA: 45% to 79%; AC1 = 0.22 to 0.71.	Not reported.

Whatman et al. (2012b)	SKB, Single leg SKB, Drop jump	Sixty six physiotherapists rated 23 young athletes (mean age 11 yr).	Video of movements. Dichotomous ratings based on pelvis position relative to neutral and knee position relative to 2nd toe. No rating instructions, ratings in real time.	Intra-rater agreement PA: 79% to 88%, AC1 = 0.60 to 0.78. Inter-rater agreement PA: 67% to 80%; AC1: 0.37 to 0.61.	Sensitivity (≥80%) and specificity (≥50%) compared to expert ratings. Peak 3D and 2D kinematics were different between groups visually rated as having good versus poor alignment by the experts.

ICC = Intraclass Correlation Coefficient, AC1 = Agreement Coefficient 1, PA = Percent Agreement, K = Kappa, SEM = Standard Error of Measure, SKB = Small Knee Bend, SLS= Single Leg Squat, LESS = Landing Error Scoring System, PFPS=Patellofemoral Pain Syndrome, FPPA = Frontal plane projection angle.

Conclusion

Visual assessment is a key skill used by clinicians in many areas of clinical practice. Visual rating of lower extremity functional tests has received attention as a promising method of identifying movement dysfunction and risk of injury. An important factor to the validity of these tests is the ability of athletes/clients to perform them in a repeatable manner. Clinicians need to be aware that for several common tests there are a variety of protocols and kinematic reliability has not been adequately assessed for many. Based on studies using 3D motion analysis, with a range of age groups and a variety of timeframes, the best evidence of reliable kinematics exists for the SLS, SKB (single or double leg) and drop jump. There is limited evidence for trunk and pelvis reliability and only one study using 3D motion analysis has reported adequate reliability of the lunge and hop lunge. Further studies are required using 3D motion analysis, looking at several tests (step-down, lunge, hop lunge) and test protocols, assessing all relevant body segments (bilaterally where possible), reporting within-day and between-days reliability and using younger populations. Additionally there is only preliminary evidence the kinematics observed during functional tests are associated with kinematics during tasks such as running and landing therefore this needs further research.

There is evidence for adequate intra-rater and inter-rater agreement for ratings of the SKB (single or double leg), SLS, single limb mini squat and drop jump across a range of ages, using a dichotomous rating of knee position relative to the foot. These dichotomous ratings have also been shown to differentiate individuals based on 2D and some 3D kinematic variables. Agreement with more complex ratings (three and four point scales, rating multiple segments) has been reported as acceptable in some studies but not others. Additionally studies have shown more complex ratings can indicate hip muscle function and 2D kinematics but they have not been compared to 3D kinematics. These more complex methods warrant further investigation as they may provide clinicians with information more relevant to making clinical decisions. For both dichotomous and more complex rating methods experience may improve intra-rater agreement and the accuracy of ratings. Accuracy is also better with slower movements. Clinicians need to be aware that additional rater training was provided in several studies reviewed and thus the reliability and validity of ratings reported may not be what is currently being achieved in clinical practice. If visual assessment is being used in isolation,

better inter-rater reliability is likely desirable for all rating methods. The relationship between 2D and 3D kinematics across a range of ages and tests needs further clarification. Several tests (e.g. lunge, hop lunge, step-down) need further investigation as the reports on these tests were few or incomplete. Many studies involved small numbers of raters limiting the generalisability of findings and thus further studies are needed with greater number of raters. Studies investigating visual assessment of clients with lower extremity dysfunction and prospective studies to determine the association between the outcome of visual screening tests and risk of injury are also needed.

Practical recommendations

When visually assessing lower extremity dynamic alignment during functional screening tests clinicians need to consider repeatability of test kinematics, intra-rater and inter-rater agreement, and rating validity. The most valid assessment in current clinical practice will be achieved with the use of a SLS or SKB (single or double leg), rating knee position relative to the foot on a dichotomous scale. If using more complex rating methods, provide example ratings and/or additional rater training. Our recommendations for clinicians currently using visual assessment, without the assistance of video slow motion are:

1. Use the SLS or SKB (single or double leg) using a protocol shown to have acceptable kinematic reliability.
2. Use a dichotomous rating of the knee position relative to the foot.
3. Use a dichotomous rating of the pelvis if wanting to assess additional segments.
4. Use experienced clinicians where possible, especially for repeat tests or if using the tests in isolation (e.g. mass screening situation).

For clinicians who have video available, slow motion video should be used to assess faster movements such as the drop jump and 2D kinematic measures may be useful.

CHAPTER 9: DISCUSSION / CONCLUSION

As stated earlier in the thesis, the assessment of movement quality/dynamic alignment during lower extremity functional screening tests is common when assessing athletes/clients for risk of injury and during rehabilitation from current injuries. Quality of movement/dynamic alignment is thought to be a key component in the management of many lower extremity pain syndromes (Sahrmann, 2002) and a lack of trunk, pelvis and lower extremity control in the frontal and transverse planes is discussed as a key risk factor for a number of injuries (Reiman, et al., 2009). Assessment of movement quality/dynamic alignment in the clinic is predominantly via visual observation, as physiotherapists do not have the time or equipment required for intricate laboratory type biomechanical analysis. However the reliability and validity of visual assessment of several lower extremity functional screening tests needs further clarification. The test protocol (e.g. variations in efforts to standardise range of hip and knee flexion, velocity of the test) (Knudson, 2000; Zeller, et al., 2003), the experience of the physiotherapist (Crossley, et al., 2011) and the rating protocol (e.g. number of scoring categories, number of segments rated, detail in rating criteria and amount of rater training) (Chmielewski, et al., 2007) may all affect the visual assessment outcome. Furthermore, there has been inadequate attention given to the repeatability of an individual's performance in several functional tests and how this performance is associated with actual function (e.g. running and landing), which are important factors in establishing them as valid clinical tools.

Given limitations in the literature, the overall question of this thesis was "Is visual and kinematic assessment of dynamic alignment during lower extremity functional screening tests reliable and valid?" Specific questions were:

I. What is the typical variation in the kinematics of adults performing SKB (double and single leg), lunge, hop lunge and step-down and young athletes performing SKB (double and single leg) and drop jump tests?

II. What is the association between kinematics during the functional tests and those during activities such as running and landing?

III. What is the reliability and validity of physiotherapist visual assessment of movement during functional tests in adults and in young athletes?

IV. What is the influence of physiotherapist experience, assessment method and test velocity on visual assessment?

V. What is the association between 2D measures of alignment and 3D kinematics during functional tests in young athletes?

The series of studies conducted to address these questions provided novel perspectives for the use of visual and kinematic assessment of movement quality/dynamic alignment during lower extremity functional screening tests. Key contributions of the thesis are provision of evidence for: (i) the reliability of 3D trunk, pelvis and lower extremity kinematics (in adults and young athletes) during a range of commonly used functional tests; (ii) functional tests being associated with actual function (running and landing); (iii) a broad range of physiotherapists being able to achieve acceptable reliability and validity with visual assessment; (iv) dichotomous ratings of knee position being the most reliable; (v) clinical experience probably improving visual assessment and (vi) slower tests being rated more accurately and reliably. The following discussion summarises the results and inferences of these studies for the main paradigms of interest: the reliability of kinematics during functional tests and their association with actual function, the reliability and validity of visual assessment of dynamic alignment/movement quality during functional tests and the influence on visual assessment of various factors including rater experience, rating method and movement velocity.

The reliability of kinematics during functional tests and their association with actual function (running and landing)

A key to the use of lower extremity functional screening tests clinically is the reliability of the participant's joint kinematics on repeat tests. Clinicians need to know the expected kinematic variability with repeat testing to aid appropriate test interpretation. Additionally as the results of the tests are often used clinically to infer kinematics during actual function (e.g. running and landing) this association waranted investigation.

In healthy adults the within-day reliability of all peak 3D trunk, pelvis and lower extremity kinematics during the descent phase of the SKB, single leg SKB, lunge, hop lunge and step-down was excellent (ICC ≥ 0.92) and the between-day reliability of the majority of kinematics was good to excellent (ICC ≥ 0.80) (chapter 4). The within- and between-day typical errors for all kinematics, across all tests, were generally small (mostly between 1° and 4°). Our methods used a mixed modelling technique as this is the most powerful for this type of analysis. This repeatability should be achievable in the clinic given the simple instructions used to standardise the range of hip and knee flexion (movement to maximum dorsiflexion while maintaining heel contact) and maintain a functional head, trunk and arm position (visual focus on a cross positioned on the wall). Standardisation of hip and knee flexion is important as the maximum frontal and transverse plane deviations need to be assessed across the same range of sagittal plane motion on any given occasion for the tests to be comparable. These results provide evidence for physiotherapists that what they are observing on any given occasion is repeatable and thus representative of a client's movement quality/dynamic alignment. Of particular clinical note was the excellent between-day reliability of peak hip kinematics (a focus of much attention in lower extremity overuse injury) and the somewhat lower between-day reliability of the trunk segment compared to all other segments. In addition to joint angles we investigated the reliability of maximum medial knee displacement (based on patella position) which showed excellent within-day (ICC ≥ 0.81) but variable between-day (ICC = 0.26 to 0.94) reliability. The variation between-day was likely due to methodological issues. Additionally the correlation between kinematics of the functional tests and jogging was generally large to very large (r = 0.53 to 0.93) providing preliminary support for the use of these tests clinically to screen kinematics during jogging gait. There are obvious clinical advantages which make the use of SKB tests more feasible than direct assessment of jogging gait. The SKB (double and single leg) and drop jump were also reliable screening tests of lower extremity dynamic alignment in young athletes (Chapter 6) – a population of particular focus in screening for risk of injury. Within-day reliability of kinematics was generally excellent (ICC ≥ 0.85) and between-day (over eight to ten weeks) was good to excellent (ICC range 0.60 to 0.92). Variability in the sagittal plane (TE: hip = 2.8° to 5.4°; knee = 2.4° to 3.7°) was greater in young athletes than in the adult population. Greater movement variability in the young athletes may have been due to variations in neuromuscular development and/or a consequence of the greater time between tests. In young athletes there was a moderate to very large (r = 0.39 to 0.87)

association between the kinematics during the small knee bend tests and those during running and landing from a drop jump. If an athlete has an injury preventing the use of higher load tests such as running and landing, an indication of dynamic alignment can still be gained via the use of small knee bend tests. Based on these findings these lower extremity functional tests could be useful in helping physiotherapists make clinical decisions regarding lower extremity movement quality/dynamic alignment and risk of injury in healthy young athlete and adult populations.

The reliability and validity of visual assessment of movement quality/dynamic alignment during functional screening tests

To our knowledge this thesis contains the first studies to report on large numbers of physiotherapists, with a range of experience and no additional visual assessment training, undertaking visual assessment (trunk, pelvis and lower extremity) of the SKB, lunge and hop lunge. For visual assessment of adults the mean intra-rater agreement was moderate to substantial (AC1: 0.39 to 0.80), range of agreement was slight to excellent (AC1: 0.01 to 0.96) and inter-rater agreement was fair to substantial (AC1: 0.22 to 0.71) (Chapter 5). This level of agreement was achieved without training of the raters but using a rating method designed to reflect current clinical practice suggesting it is achievable in the clinic. The substantial variation in intra-rater agreement between some physiotherapists confirms previous studies with small numbers of physiotherapists may not be generalisable to all physiotherapists. Substantial intra-rater agreement (AC1: 0.60 to 0.78) was also achieved by most physiotherapists visually rating the pelvis and knee, using a dichotomous scale, in young athletes during SKB and drop jump tests (Chapter 7). A dichotomous rating of only the pelvis and knee was used with the young athletes as these were the most reliable in adults and in keeping with protocols used by other authors Inter-rater agreement was moderate to substantial (AC1: 0.52 to 0.61) for the SKB (double and single leg) but only fair for the drop jump.. Compared to expert consensus ratings (aided by video pause and rewind) the majority of physiotherapists achieved acceptable sensitivity (≥80%) and specificity (≥50%) for the pelvis and knee ratings of young athletes performing the SKB tests but only acceptable specificity for the drop jump. The expert consensus ratings were able to differentiate young athletes with different 2D kinematics (very likely to almost

certainly) and 3D hip kinematics (likely to very likely) but not 3D knee abduction. In the young athletes 2D kinematic measures were moderately associated (r = 0.33 to 0.47) with 3D hip kinematics, but not 3D knee kinematics and strongly associated (0.71 to 0.88) with pelvic, femoral and tibial segment rotations in the frontal plane. Although there are time and resource implications simple 2D measures may be a useful addition to clinical assessment. However, 2D measures should not be considered a substitute for quantification of 3D joint rotations.

The influence on visual assessment of various factors including rater experience, rating method and movement velocity

To our knowledge this thesis contains the first studies to quantify differences in assessment as a result of clinical experience, rating method and movement velocity. There was some evidence that more experienced physiotherapists reached higher levels of intra-rater agreement when rating adults and this was likely to be clinically important (84 to 99% likelihood) but this was less likely for inter-rater agreement (Chapter 5). Experience did not improve agreement with visual assessment of young athletes (Chapter 7), although rating accuracy (compared to expert consensus ratings) did improve with experience for some but not all ratings (DOR 1.6 to 2.8 times better). This variable influence of experience may reflect rating movement from video was new to the majority of physiotherapists and experience with this method was similar across the group. Alternatively success with visual rating may be more a consequence of the perceptual style of the observer and/or rating method (Morrison, 2000) than clinical experience.

For ratings of adults, intra-rater and inter-rater agreement using a dichotomous scale was better for all physiotherapists than using an ordinal four point scale and this difference was very likely clinically important (97 to 100% likelihood) (Chapter 5). Rating of the knee position relative to the foot achieved the highest intra- and inter-rater agreement of all the segments rated. However, using a segmental versus overall rating method did not consistently influence the level of agreement achieved. It may be that visual rating can be done using an overall approach and clinically this is probably not uncommon, however this requires further research.

In the visual assessment of young athletes intra-rater agreement was better for the slower SKB tests than for the faster drop jump and the accuracy of ratings was also improved (DOR 4.9 times better) (Chapter 7). This suggests that if sufficient information can be gained from the SKB it is the more valid clinical tool. The use of video slow motion (and or simple 2D measures of alignment) may need to be considered to achieve more accurate ratings of the drop jump, although this has resource implications for clinical use.

Thesis limitations

Methodological limitations must be considered in relation to this thesis research and were discussed where relevant in each chapter. In summary, these limitations mainly surrounded the use of healthy populations, a small sample of adults rated and video based visual ratings.

- Repeated tests in adults were over a one to two day time period and this may not represent typical kinematic variations over longer timeframes.
- We investigated the reliability of kinematics with uninjured populations only.
- Sample sizes used to assess the association between kinematics during the functional tests and those during running and landing were relatively small and thus caution is needed for interpretation of some correlations reported.
- The sample of adults visually rated was relatively homogenous and small in order to keep the time required to complete the ratings to an acceptable length.
- Ratings of the adults and young athletes were only made from an anterior view as this appeared common in clinical screening of these functional tests.
- All visual ratings were made from video recordings as we wanted all physiotherapists to see exactly the same movement and it was not practical to have all physiotherapists observing tests in real time.
- The presence of anatomical markers (required for the 3D kinematic analysis) may have influenced visual ratings.

- For logistical reasons the environment for the experienced physiotherapists (when rating the adults) was different to that for the inexperienced and novice physiotherapists.

While this thesis reports substantial new findings that further broaden the body of knowledge surrounding lower extremity functional screening tests, their kinematic reliability and association with actual function, the reliability and validity of physiotherapist visual ratings and the influence of experience, ratings method and movement velocity, the results need to be interpreted with caution given the aforementioned thesis limitations.

Recommendations for future research

The findings of this thesis have lead to the following recommendations for future research:

- Populations with lower extremity injury, such as patellofemoral pain, should be a focus of future research looking at both kinematic and visual assessment of lower extremity functional screening tests. Such studies should investigate the use of these tests in identifying readiness for return to sport and progress during rehabilitation.

- Prospective studies are required to investigate the ability of lower extremity functional screening tests to predict risk of injury.

- The reliability of kinematics in adults over longer periods of time warants investigation and studies investigating the validity of visual ratings of the lunge and hop lunge are needed.

- Studies of larger populations would increase the precision in estimating the association between functional test kinematics and kinematics during actual function such as running and landing.

- Ratings made in real time (rather than from recorded video) may give a better indication of agreement in clinical practice, so should be investigated.

- For faster movements such as the Drop Jump the use of video slow motion and pause/rewind may improve rating accuracy. A study investigating the velocity of movement viewed via different media should be conducted.

- The ability of visual ratings to differentiate athletes with kinetics that increase risk fo injury and the influence of markers on visual rating warants investigation.

- An intervention study using kinematic and visual assessment of the SKB as an outcome measure to determine the ability of physiotherapists to change movement dysfuntion should be conducted.

Conclusion

In healthy adults and young athletes the reliability of peak 3D kinematics during the descent phase of common lower extremity functional screening tests (using our simple instructions) is acceptable for the majority of kinematic variables of interest to physiotherapists. Physiotherapists can thus be confident that what they observe in these tests is representative of the client/athletes movement quality/dynamic alignment. Additionally there is a moderate to large association between peak functional test kinematics and those recorded during running and landing. Therefore these easily administered clinical tests, without the need for specialised equipment or additional space, can give the physiotherapist an indication of the movement pattern/dynamic alignment that is likely during higher load activities often associated with injury. Based on these results lower extremity functional screening tests should be useful in helping physiotherapists make clinical decisions regarding trunk, pelvis and lower extremity movement quality/dynamic alignment and risk of injury.

A broad range of physiotherapists, without the need for additional training, achieved substantial agreement (intra-rater and inter-rater) when visually rating dynamic knee and pelvis position (on a dichotomous scale), during functional screening tests in young athletes. Intra-rater agreement was better for the SKB tests than the faster drop jump. Additionally physiotherapists' visual rating of movement quality/dynamic alignment in adults during a range of lower extremity functional tests (using dichotomous and ordinal scales) resulted in moderate to substantial mean intra-rater and fair to substantial inter-rater agreement. The best agreement was achieved by the dichotomously classified ratings of experienced physiotherapists, irrespective of whether a segment or overall rating approach was used. In adults, knee position relative to the foot achieved the best agreement of all segments rated. Increased experience was likely to result in a clinically important improvement in intra-rater agreement (adult ratings only) but not inter-rater agreement.

Additionally sensitivity and specificity was acceptable for all ratings of young athletes performing the small knee bend tests when compared to the video assisted consensus ratings of three experts. The faster Drop Jump test achieved acceptable specificity but not sensitivity and was not rated as accurately as the slower SKB. Increased experience may improve the accuracy of ratings in young athletes but this was not consistent across all ratings. For all tests the expert consensus ratings were able to differentiate young athletes with different hip and pelvis kinematics that may require intervention.

As visual assessments are one of many used in making a clinical decision we suggest the levels of agreement reached across all the functional tests investigated (particularly by the experienced physiotherapists using the dichotomous scale) are acceptable for clinical use. However better levels of inter-rater agreement would likely be desirable for use of the tests independently in a mass screening situation. Visual rating of movement quality/dynamic alignment by physiotherapists during lower extremity functional screening tests should provide reliable and valid information that assists in the clinical decision making process.

For physiotherapists in clinical practice wanting to use a visual rating method supported by the best current evidence we would recommend use of the SKB (double or single leg) or SLS, using a protocol with acceptable kinematic reliability and a dichotomous scale to rate the position of the knee and/or pelvis. This is the most valid method for visually rating the dynamic lower extremity alignment of healthy individuals in the clinic, without the use of video pause and rewind.

Additional recommendations include:
- As first choice use the SKB (double and/or single leg), with our simple instructions to maintain a functional head, trunk and arm position and standardise hip and knee flexion range of motion.
- If the assessment requires additional tests use a lunge or hop lunge.
- Use video slow motion and/or pause if assessing the drop jump.
- Focus on dichotomous ratings of knee and/or pelvis position.
- Where possible repeat tests should be performed by the same physiotherapist.
- Where possible use physiotherapists with the most clinical experience.

REFERENCES

Adirim, T. A., & Cheng, T. L. (2003). Overview of injuries in the young athlete. *Sports Medicine, 33*(1), 75-81.

Ageberg, E., Bennell, K. L., Hunt, M. A., Simic, M., Roos, E. M., & Creaby, M. W. (2010). Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat. *BMC Musculoskeletal Disorders, 11*, 265. doi:10.1186/1471-2474-11-265

Alkjær, T., Henriksen, M., Dyhre-Poulsen, P., & Simonsen, E. B. (2009). Forward lunge as a functional performance test in ACL deficient subjects: Test-retest reliability. *The Knee, 16*(3), 176-182.

Angeloni, C., Cappozzo, A., Catani, F., & Leardini, A. (1993). Quantification of relative displacement of skin- and plate- mounted markers with respect to bones. *Journal of Biomechanics, 26*, 864.

Bahr, R., & Holme, I. (2003). Risk factors for sports injuries - a methodological approach. *British Journal of Sports Medicine, 37*(5), 384-392.

Baker, R. (2001). Pelvic angles: a mathematically rigorous definition which is consistent with a conventional clinical understanding of the terms. *Gait & Posture, 13*, 1-6.

Bell, D. R., Padua, D. A., & Clark, M. A. (2008). Muscle strength and flexibility characteristics of people displaying excessive medial knee displacement. *Archives of Physical Medicine & Rehabilitation, 89*(7), 1323-1328.

Bernhardt, J., Bate, P. J., & Matyas, T. A. (1998). Accuracy of observational kinematic assessment of upper-limb movements. *Physical Therapy, 78*(3), 259-270.

Blood, E., & Spratt, K. (2007, April 16-19). Disagreement on Agreement: Two Alternative Agreement Coefficients Symposium conducted at the meeting of the SAS Global Forum 2007, Statistics and Data Analysis, Florida.

Bolgla, L. A., & Keskula, D. R. (1997). Reliability of lower extremity functional performance tests. *Journal of Orthopaedic and Sports Physical Therapy, 26*(3), 138-142.

Bolgla, L. A., Malone, T. R., Umberger, B. R., & Uhl, T. L. (2008). Hip strength and hip and knee kinematics during stair descent in females with and without patellofemoral pain syndrome. *Journal of Orthopaedic and Sports Physical Therapy, 38*(1), 12-18.

Boling, M. C., Padua, D. A., Marshall, S. W., Guskiewicz, K., Pyne, S., & Beutler, A. (2009). A Prospective Investigation of Biomechanical Risk Factors for Patellofemoral Pain Syndrome. *The American Journal of Sports Medicine, 37*(11), 2108-2116. doi:10.1177/0363546509337934

Brunnekreef, J. J., van Uden, C. J., van Moorsel, S., & Kooloos, J. G. (2005). Reliability of videotaped observational gait analysis in patients with orthopedic impairments. *BMC Musculoskeletal Disorders, 6*(1), 17.

Caine, D., Caine, C., & Maffulli, N. (2006). Incidence and Distribution of Pediatric Sport-Related Injuries. *Clinical Journal of Sport Medicine, 16*(6), 500-513.

Caine, D., Maffulli, N., & Caine, C. (2008). Epidemiology of injury in child and adolescent sports: injury rates, risk factors, and prevention. *Clinics in Sports Medicine, 27*(1), 19-50.

Caine, D. J., & Golightly, Y. M. (2011). Osteoarthritis as an outcome of paediatric sport: an epidemiological perspective. *British Journal of Sports Medicine, 45*(4), 298-303.

Cappozzo, A., Cappello, A., Della Croce, U., & Pensalfini, F. (1997). Surface marker cluster design for 3-D bone movement reconstruction. *IEEE Transactions on Biomedical Engineering, 44*(12), 1165-1174.

Chan, Y. (2003). Biostatistics 104: Correlational Analysis. *Singapore Medical Journal, 44*(12), 614-619.

Cheng, H., Laron, M., Schiffman, J. S., Tang, R. A., & Frishman, L. J. (2007). The relationship between visual field and retinal nerve fiber layer measurements in patients with multiple sclerosis. *Investigative Ophthalmology and Visual Science, 48*(12), 5798-5805.

Chiaia, T. A., Maschi, R. A., Stuhr, R. M., Rogers, J. R., Sheridan, M. A., Callahan, L. R., et al. (2009). A musculoskeletal profile of elite female soccer players. *HSS Journal, 5*(2), 186-195.

Chmielewski, T. L., Hodges, M. J., Horodyski, M., Bishop, M. D., Conrad, B. P., & Tillman, 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 and Sports Physical Therapy, 37*(3), 122-129.

Claiborne, T. L., Armstrong, C. W., Gandhi, V., Pincivero, D. M., Claiborne, T. L., Armstrong, C. W., et al. (2006). Relationship between hip and knee strength and knee valgus during a single leg squat. *Journal of Applied Biomechanics, 22*(1), 41-50.

Clark, N. (2001). Functional performance testing following knee ligament injury. *Physical Therapy in Sport, 2*(2), 91-105.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*(1), 37-46.

Collins, N. J., Bisset, L. M., Crossley, K. M., & Vicenzino, B. (2012). Efficacy of Nonsurgical Interventions for Anterior Knee Pain: Systematic Review and Meta-Analysis of Randomized Trials. *Sports Medicine, 42*(1), 31-49 10.2165/11594460-000000000-000000000.

Cook, G., Burton, L., & Hoogenboom, B. (2006). Pre-participation screening: The use of fundamental movements as an assessment of function - part 1. *North American journal of sports physical therapy, 1*(2), 62-72.

Cook, J. (2006). Patella tendon rehabilitation. In P. Brukner & K. Khan (Eds.), *Clinical Sports Medicine DVD*. Sydney: McGraw. Retrieved from http://www.clinicalsportsmedicine.com/dvd.htm

Cowan, S. M., Crossley, K. M., & Bennell, K. L. (2008). Altered hip and trunk muscle function in individuals with patellofemoral pain. *British Journal of Sports Medicine, 43*(8), 584-588.

Crossley, K., Bryant, A., Jing Zhang, W., & Cowan, S. (2006). A simple clinical assessment tool may predict abnormal hip muscle function. *Journal of Science and Medicine in Sport, 9*(Supplement 1), 6-7.

Crossley, K., Cook, J., Cowan, S., & McConnell, J. (2006). Anterior Knee Pain. In P. Brukner & K. Khan (Eds.), *Cinical Sports Medicine* (3rd ed.). Sydney: McGraw-Hill.

Crossley, K. M., Zhang, W. J., Schache, A. G., Bryant, A., & Cowan, S. M. (2011). Performance on the single-leg squat task indicates hip abductor muscle function. *American Journal of Sports Medicine, 39*(4), 866-873.

Di Eugenio, B., & Glass, M. (2004). The kappa statistic: a second look. *Computational Linguistics, 30*(1), 95-101.

DiMattia, M., Livengood, A., Uhl, T., Mattacola, C., & Malone, T. (2005). What are the validity of the single-leg-squat test and its relationship to hip-abduction strength? *Journal of Sport Rehabilitation, 14*, 108-123

Dwyer, M., Boudreau, S., Mattacola, C., Uhl, T., & Lattermann, C. (2010). Comparison of Lower Extremity Kinematics and Hip Muscle Activation During Rehabilitation Tasks Between Sexes. *Journal of Athletic Training, 45*(2), 181-190.

Earl, J. E., Hertel, J., & Denegar, C. R. (2005). Patterns of dynamic malalignment, muscle activation, joint motion, and patellofemoral-pain syndrome. *Journal of Sport Rehabilitation, 14*, 215-233

Earl, J. E., Monteiro, S. K., & Snyder, K. R. (2007). Differences in lower extremity kinematics between a bilateral drop-vertical jump and a single-leg step-down. *Journal of Orthopaedic and Sports Physical Therapy, 37*(5), 245-252.

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 and Sports Physical Therapy, 39*(9), 665-674. doi:10.2519/jospt.2009.3004

Emery, C. A. (2010). Injury prevention in paediatric sport-related injuries: a scientific approach. *British Journal of Sports Medicine, 44*(1), 64-69. doi:10.1136/bjsm.2009.068353

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

Fleiss, J. L. (1999). *The Design and Analysis of Clinical Experiments* New York: Wiley.

Ford, K. R., Myer, G. D., & Hewett, T. E. (2007). Reliability of landing 3D motion analysis: implications for longitudinal analyses. *Medicine and Science in Sports and Exercise, 39*(11), 2021-2028. doi:10.1249/mss.0b013e318149332d

Ford, K. R., Myer, G. D., Smith, R. L., Vianello, R. M., Seiwert, S. L., & Hewett, T. E. (2006). A comparison of dynamic coronal plane excursion between matched male and female athletes when performing single leg landings. *Clinical Biomechanics, 21*(1), 33-40.

Fritz, J. M., & Wainner, R. S. (2001). Examining diagnostic tests: an evidence-based perspective. *Physical Therapy 81*(9), 1546-1564.

Frohm, A., Heijne, A., Kowalski, J., Svensson, P., & Myklebust, G. (2011). A nine-test screening battery for athletes: a reliability study. *Scandinavian Journal of Medicine & Science in Sports*. doi:10.1111/j.1600-0838.2010.01267.

Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. (2003). The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology, 56*(11), 1129-1135.

Grood, E. S., & Suntay, W. J. (1983). A joint coordinate system for the clinical description of three-dimensional motions: application to the knee. *Journal of Biomechanical Engineering, 105*(2), 136-144.

Gwet, K. (2001). *Handbook of Inter-rater Reliability: How to estimate the level of agreement between two or multiple raters.* Gaithersburg: STATAXIS Publishing Company.

Gwet, K. (2002). *Inter-Rater Reliability: Dependency on Trait Prevalence and Marginal Homogeneity*. Statistical Methods For Inter-Rater Reliability Assessment No.2. Retrieved from http://agreestat.com/research_papers.html

Gwet, K. L. (2008a). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical & Statistical Psychology, 61*(1), 29.

Gwet, K. L. (2008b). Variance Estimation of Nominal-Scale Inter-Rater Reliability with Random Selection of Raters. *Psychometrika, 73*(3), 407-407-430. doi:10.1007/s11336-007-9054-8

Haley, D. (2007). *Using a New Inter-rater Reliability Statistic* (PhD dissertion). The Open University, Milton Keynes.

Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures, 1*(1), 77-89. doi:10.1080/19312450709336664

Herrington, L. (2009). Knee valgus angle during landing tasks in female volleyball and basketball players. *Journal of Strength and Conditioning Research, 25*(1), 262-266. doi:10.1519/JSC.0b013e3181b62c77

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

Hirth, C. J., & Padua, D. A. (2007). Clinical Movement Analysis to Identify Muscle Imbalances and Guide Exercise. *Athletic Therapy Today, 12*(4), 10-14.

Hoehler, F. K. (2000). Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *Journal of Clinical Epidemiology, 53*(5), 499-503.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine, 30*(1), 1-15.

Hopkins, W. G. (2006). A spreadsheet for combining outcomes from several subject groups. *Sportscience, 10*, 51-53.

Hopkins, W. G. (2007). A spreadhseet to compare means of two groups. *Sportscience 11*, 22-23.

Hopkins, W. G., & Manly, B. F. (1989). Errors in assigning grades based on tests of finite validity. *Research quaterly for exercise and sport, 60*(2), 180-182.

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise, 41*(1), 3-13. doi:10.1249/MSS.0b013e31818cb278

Ireland, M. L., Willson, J. D., Ballantyne, B. T., & Davis, I. M. (2003). Hip strength in females with and without patellofemoral pain. *Journal of Orthopaedic & Sports Physical Therapy, 33*(11), 671-676.

Kadaba, M. P., Ramakrishnan, H. K., Wootten, M. E., Gainey, J., Gorton, G., & Cochran, G. V. (1989). Repeatability of kinematic, kinetic, and electromyographic data in normal adult gait. *Journal of Orthopaedic Research, 7*(6), 849-860. doi:10.1002/jor.1100070611

Kernaghan, D., & Penney, G. (2006). Do panels vary when assessing intrapartum adverse events? The reproducibility of assessments by hospital risk management groups. *Qaulity and Safety in Health Care, 15*, 359-362.

Kernozek, T. W., Torry, M. R., van Hoof, H., Cowley, H., & Tanner, S. (2005). Gender differences in frontal and sagittal plane biomechanics during drop landings. *Medicine and Science in Sports and Exercise, 37*(6), 1003-1012.

Kibler, W. B., Uhl, T. L., Maddux, J. W. Q., Brooks, P. V., Zeller, B., & McMullen, J. (2002). Qualitative clinical evaluation of scapular dysfunction: A reliability study. *Journal of Shoulder and Elbow Surgery, 11*(6), 550-556.

Knudson, D. (1999). Validity and reliability of visual ratings of the vertical jump. *Perceptual and Motor Skills, 89*, 642-648.

Knudson, D. (2000). What can professionals qualitatively analyze? *Journal of Physical Education, Recreation & Dance, 71*(2), 19-19-23.

Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B. J., HrÃ³bjartsson, A. r., et al. (2011). Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *International Journal of Nursing Studies, 48*(6), 661-671.

Kritz, M. M. C., Cronin, J. P., & Hume, P. P. (2009). Using the Body Weight Forward Lunge to Screen an Athlete's Lunge Pattern. *Strength & Conditioning Journal, 31*(6), 15-24.

Krosshaug, T., Nakamae, A., Boden, B., Engebretsen, L., Smith, G., Slauterbeck, J., et al. (2007). Estimating 3D joint kinematics from video sequences of running and cutting maneuvers-- assessing the accuracy of simple visual inspection. *Gait & Posture, 26*(3), 378-385.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Leetun, D. T., Ireland, M. L., Willson, J. D., Ballantyne, B. T., & Davis, I. M. (2004). Core stability measures as risk factors for lower extremity injury in athletes. *Medicine and science in sports and exercise, 36*(6), 926-934.

Levinger, P., Gilleard, W., & Coleman, C. (2007). Femoral medial deviation angle during a one-leg squat test in individuals with patellofemoral pain syndrome. *Physical Therapy in sport, 8*(4), 163-168.

Lord, S. E., Halligan, P. W., & Wade, D. T. (1998). Visual gait analysis: The development of a clinical assessment and scale. *Clinical Rehabilitation, 12*(2), 107-119.

Manal, K., McClay, I., Stanhope, S., Richards, J., & Galinat, B. (2000). Comparison of surface mounted markers and attachment methods in estimating tibial rotations during walking: an in vivo study. *Gait & Posture, 11*(1), 38-45.

Mascal, C. L., Landel, R., & Powers, C. (2003). Management of patellofemoral pain targeting hip, pelvis, and trunk muscle function: 2 case reports. *Journal of Orthopaedic and Sports Physical Therapy, 33*(11), 647-660.

McLean, S. G., Walker, K., Ford, K. R., Myer, G. D., Hewett, T. E., & van den Bogert, A. J. (2005). Evaluation of a two dimensional analysis method as a screening and evaluation tool for anterior cruciate ligament injury. *British Journal of Sports Medicine, 39*(6), 355-362.

McLeod, T. C. V., Decoster, L. C., Loud, K. J., Micheli, L. J., Parker, J. T., Sandrey, M. A., et al. (2011). National Athletic Trainers' Association position statement: prevention of pediatric overuse injuries. *Journal of Athletic Training, 46*(2), 206(215).

Morrison, C. S. (2000). Qualitative analysis: Putting it all together: Why don't you analyze the way I analyze? *Journal of Physical Education, Recreation & Dance, 71*(1), 22-22-25.

Mottram, S., & Comerford, M. (2008). A new perspective on risk assessment. *Physical Therapy in Sport, 9*(1), 40-51.

Noyes, F. R., Barber-Westin, S. D., Fleckenstein, C., Walsh, C., & West, J. (2005). The drop-jump screening test: Difference in lower limb control by gender and effect of neuromuscular training in female athletes. *American Journal of Sports Medicine, 33*(2), 197-207.

Olson, T. J., Chebny, C., Willson, J. D., Kernozek, T. W., & Straker, J. S. (2011). Comparison of 2D and 3D kinematic changes during a single leg step down following neuromuscular training. *Physical Therapy in Sport, 12*(2), 93-99.

Onate, J., Cortes, N., Welch, C., & van Lunen, B. (2010). Expert Versus Novice Interrater Reliability and Criterion Validity of the Landing Error Scoring System. *Journal of Sport Rehabilitation, 19*(1), 41-56.

Örtqvist, M., Moström, E., Roos, E., Lundell, P., Janarv, P.-M., Werner, S., et al. (2011). Reliability and reference values of two clinical measurements of dynamic and static knee position in healthy children. *Knee Surgery, Sports Traumatology, Arthroscopy*, 1-7. doi:10.1007/s00167-011-1542-9

Padua, D. A., Marshall, S. W., Boling, M. C., Thigpen, C. A., Garrett, W. E., Jr., & 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. *American Journal of Sports Medicine, 37*(10), 1996-2002. doi:10.1177/0363546509343200

Pantano, K. J., White, S. C., Gilchrist, L. A., & Leddy, J. (2005). Differences in peak knee valgus angles between individuals with high and low Q-angles during a single limb squat. *Clinical Biomechanics, 20*(9), 966-972.

Passier, L. N., Nasciemento, M. P., Gesch, J. M., & Haines, T. P. (2010). Physiotherapist observation of head and neck alignment. *Physiotherapy Theory and Practice, 26*(6), 416-423. doi:10.3109/09593980903317557

Piva, S. R., Fitzgerald, K., Irrgang, J. J., Jones, S., Hando, B. R., Browder, D. A., et al. (2006). Reliability of measures of impairments associated with patellofemoral pain syndrome. *BMC Musculoskeletal Disorders. 7:33.*

Powers, C. M. (2003). The influence of altered lower-extremity kinematics on patellofemoral joint dysfunction: A theoretical perspective. *Journal of Orthopaedic and Sports Physical Therapy, 33*(11), 639-646.

Powers, C. M. (2010). The influence of abnormal hip mechanics on knee injury: a biomechanical perspective. *Journal of Orthopaedic & Sports Physical Therapy, 40*(2), 42-51. doi:10.2519/jospt.2010.3337

Reid, D., Stotter, G., Schneiders, T., Hing, W., & White, G. (2003). *NZ Academy of Sport Musculoskeletal Screening Manual*. Auckland

Reiman, M. P., Bolgla, L. A., & Lorenz, D. (2009). Hip Function's Influence on Knee Dysfunction: A Proximal Link to a Distal Problem. *Journal of Sport Rehabilitation, 18*(1), 33-46.

Russell, K. A., Palmieri, R. M., Zinder, S. M., & Ingersoll, C. D. (2006). Sex Differences in Valgus Knee Angle During a Single-Leg Drop Jump. *Journal of Athletic Training, 41*(2), 166-171.

Sahrmann, S. A. (1998). *Diagnosis and Treatment of Movement Impairment Syndromes - Course Notes*. Course notes. London.

Sahrmann, S. A. (2002). *Diagnosis and Treatment of Movement Impairment Syndromes*. London: Mosby.

Sahrmann, S. A. (2011). On "The physical therapy and society summit (PASS) meeting..." Kigin CM, Rodgers MM, Wolf SL; for the PASS steering committee members. Phys Ther. 2010;90:1555-1567. *Physical Therapy, 91*(3), 432-433; author reply 433-435. doi:10.2522/ptj.2011.91.3.432

Sheerin, K., Whatman, C., Hume, P. A., & Croft, J. (2010. *Reliability of 3D gait kinematics in treadmill running in healthy young athletes*. Paper presented at the meeting of the International Society of Biomechanics in Sports 2010 Congress, Marquette, U.S.A.

Sigward, S. M., Ota, S., & Powers, C. M. (2008). Predictors of frontal plane knee excursion during a drop land in young female soccer players. *Journal of Orthopaedic and Sports Physical Therapy, 38*(11), 661-667. doi:10.2519/jospt.2008.2695

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

Sole, G., Milosavljevic, S., Sullivan, S. J., & Nicholson, H. (2008). Running-related hamstring injuries: a neuromuscular approach. *Physical Therapy Reviews, 13*(2), 102-110.

Somers, D. L., Hanson, J. A., Kedzierski, C. M., Nestor, K. L., & Quinlivan, K. Y. (1997). The influence of experience on the reliability of goniometric and visual measurement of forefoot position. *Journal of Orthopaedic and Sports Physical Therapy, 25*(3), 192-202.

Souza, R. B., & Powers, C. M. (2009). Differences in hip kinematics, muscle strength, and muscle activation between subjects with and without patellofemoral pain. *Journal of Orthopaedic & Sports Physical Therapy, 39*(1), 12-19.

Stensrud, S., Myklebust, G., Kristianslund, E., Bahr, R., & Krosshaug, T. (2011). Correlation between two-dimensional video analysis and subjective assessment in evaluating knee control among elite female team handball players. *British Journal of Sports Medicine, 45*(7), 589-595. doi:10.1136/bjsm.2010.078287

Thijs, Y., Van Tiggelen, D., Willems, T., De Clercq, D., & Witvrouw, E. (2007). Relationship between hip strength and frontal plane posture of the knee during a forward lunge. *British Journal of Sports Medicine, 41*(11), 723-727. doi:10.1136/bjsm.2007.037374

Tofte KB, Tillman SM, & Chmielewski TL. (2011). Functional movement analysis can identify differences in frontal plane projection angle in female athletes performing heel tap. *Journal of Orthopaedic & Sports Physical Therapy, 41*(1), A39-A40.

Trulsson, A., Garwicz, M., & Ageberg, E. (2010). Postural orientation in subjects with anterior cruciate ligament injury: development and first evaluation of a new observational test battery. *Knee Surgery Sports Traumatology Arthroscopy, 18*(6), 814-823. doi:10.1007/s00167-009-0959-x

Uebersax, J. (n.d.). *Statitstical methods for rater and diagnostic agreement*. Retrieved September 12, 2011, from http://www.john-uebersax.com/stat/agree.htm

Vach, W. (2005). The dependence of Cohen's kappa on the prevalence does not matter. *Journal of Clinical Epidemiology, 58*(7), 655-661. doi:10.1016/j.jclinepi.2004.02.021

von Porat, A., Holmstrom, E., & Roos, E. (2008). Reliability and validity of videotaped functional performance tests in ACL-injured subjects. *Physiotherapy Research International, 13*(2), 119-130.

Weir, A., Darby, J., Inklaar, H., Koes, B., Bakker, E., & Tol, J. L. (2010). Core stability: inter- and intraobserver reliability of 6 clinical tests. *Clinical Journal of Sport Medicine, 20*(1), 34-38.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research, 19*(1), 231-240. doi:10.1519/15184.1

Whatman, C., Hing, W., & Hume, P. (2011a). Kinematics during lower extremity functional screening tests-Are they reliable and related to jogging? *Physical Therapy in sport, 12*(1), 22-29.

Whatman, C., Hing, W., & Hume, P. A. (2011b). Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. *Physical Therapy in Sport, in press*. doi:10.1016/j.ptsp.2011.07.001

Whatman, C., Hume, P. A., & Hing, W. (2012a). Kinematics during lower extremity functional screening tests in young athletes – are they reliable and valid? *Physical Therapy in sport, In Review*.

Whatman, C., Hume, P. A., & Hing, W. (2012b). The reliability and validity of physiotherapist visual rating of dynamic pelvis and knee alignment in young athletes. *Physical Therapy in sport, In Review*.

Whatman, C., Hume, P. A., & Hing, W. (2012c). The associations between 2D alignment and 3D kinematics during lower extremity functional tests in a group of young athletes. *Sport Biomechanics, In Review*.

Willems, T. M., De Clercq, D., Delbaere, K., Vanderstraeten, G., De Cock, A., & Witvrouw, E. (2006). A prospective study of gait related risk factors for exercise-related lower leg pain. *Gait Posture, 23*(1), 91-98.

Willson, J. D., & Davis, I. S. (2008a). Lower extremity mechanics of females with and without patellofemoral pain across activities with progressively greater task demands. *Clinical Biomechanics, 23*(2), 203-211.

Willson, J. D., & Davis, I. S. (2008b). Utility of the Frontal Plane Projection Angle in Females with Patellofemoral Pain. *Journal of Orthopaedic & Sports Physical Therapy, 38*(10), 606-616.

Willson, J. D., & Davis, I. S. (2009). Lower extremity strength and mechanics during jumping in women with patellofemoral pain. *Journal of Sport Rehabilitation, 18*(1), 76-90.

Willson, J. D., Ireland, M. L., & Davis, I. M. (2006). Core strength and lower extremity alignment during single leg squats. *Medicine and Science in Sports and Exercise, 38*(5), 945-952.

Zazulak, B. T., Hewett, T. E., Reeves, N. P., Goldberg, B., & Cholewicki, J. (2007). Deficits in neuromuscular control of the trunk predict knee injury risk: a prospective biomechanical-epidemiologic study. *American Journal of Sports Medicine, 35*(7), 1123-1130. doi:10.1177/0363546507301585

Zeller, B. L., McCrory, J. L., Kibler, W. B., & Uhl, T. L. (2003). Differences in kinematics and electromyographic activity between men and women during the single-legged squat. *American Journal of Sports Medicine, 31*(3), 449-456.

APPENDIX 1: PHYSIOTHERAPIST AGREEMENT IN RATING MOVEMENT QUALITY DURING LOWER EXTREMITY FUNCTIONAL TESTS

Whatman, C., Hing, W., & Hume, P. A. (2008, November). Physiotherapist agreement in rating movement quality during lower extremity functional tests. *Sports Medicine New Zealand.* Symposium conducted at the meeting of the New Zealand Conference of Medicine and Science in Sport, Dunedin.

PHYSIOTHERAPIST AGREEMENT IN RATING MOVEMENT QUALITY DURING LOWER EXTREMITY FUNCTIONAL TESTS

Whatman C[1,2], Hing W[1], Hume P[2]

[1]Health and Rehabilitation Research Centre, AUT University, Auckland, New Zealand, [2]Institute of Sport and Recreation Research New Zealand (ISRRNZ), AUT University, Auckland, New Zealand

Purpose: Visual observation of movement quality is commonly used in clinical practice during rehabilitation or when performing musculoskeletal screening for risk of future injury. The purpose of this study was to determine physiotherapist agreement (intrarater and interrater) for two methods of rating movement quality during four lower extremity functional tests.

Methods: Video recordings were captured of seven uninjured participants performing four lower extremity functional tests (small knee bend (SKB), single SKB, lunge and single leg land). These recordings were viewed by 10 expert physiotherapists on two separate occasions. On each occasion all physiotherapists visually rated overall movement quality as well as movement quality of the trunk, pelvis, knee and foot separately. Ratings for each of these segments were based on a judgement as to whether participants maintained an acceptable segment position throughout each test. Descriptions of acceptable segment position were based on common clinical practice. All ratings were recorded on a standardised rating scale specifically designed for this study. This scale included a yes/no rating to indicate if the movement pattern/segment position was acceptable and a three point scale (minor, moderate, marked) to score the severity of any unacceptable ratings. To determine the level of agreement between physiotherapists overall percent agreement was calculated for the yes/no ratings and ratings on the three point scale.

Results: For yes/no ratings, combined for each of the four tests, mean intrarater percent agreement (80% to 85%) was higher than interrater percent agreement (64% to 75%). This was also the case when ratings on the three point scale were included with mean intrarater percent agreement (56% to 79%) and interrater percent agreement (35% to 64%). Further analysis, including appropriate coefficients of agreement will be presented.

Conclusions: Based on these preliminary results it appears physiotherapists demonstrate good intrarater and moderate interrater agreement when visually rating movement quality during lower limb functional tests. Agreement on a yes/no scale is better than agreement on a scale that includes minor, moderate and marked.

APPENDIX 2: LOWER EXTREMITY FUNCTIONAL TESTS – ARE THEY RELIABLE AND DO THE PREDICT ACTUAL FUNCTION?

Whatman, C., Hume, P. A., & Hing, W. (2009, November). Lower extremity functional tests – are they reliable and do they predict actual function? *Sports Medicine New Zealand.* Symposium conducted at the meeting of the New Zealand Conference of Medicine and Science in Sport, Rotorua.

LOWER EXTREMITY FUNCTIONAL TESTS – ARE THEY RELIABLE AND DO THEY PREDICT ACTUAL FUNCTION?

Whatman C[1,2], Hume P[2], Hing W[1],

[1]Health and Rehabilitation Research Centre, AUT University, Auckland, New Zealand, [2]Institute of Sport and Recreation Research New Zealand (ISRRNZ), AUT University, Auckland, New Zealand

Purpose: To investigate the within-day and between-day reliability of 3D kinematics during five lower extremity functional tests and to assess the relationship between these kinematics and those recorded during jogging.

Methods: Twenty five uninjured participants attended the AUT University Motion Analysis Laboratory on one occasion to perform five lower extremity functional tests (see Table 1). Ten participants returned on a second occasion for repeat testing. A nine camera motion analysis system capable of recording whole body 3D kinematics was utilised. Prior to data collection the trunk, pelvis and dominant leg of all participants were instrumented with 15 retro-reflective markers secured to specific anatomical locations. A demonstration of each movement and standardized verbal instructions were given prior to each of the five functional tests performed by all participants (see Table 1). Participants performed three repetitions of each test and were also recorded jogging. For all kinematics of interest (trunk, pelvis, hip, knee, foot) typical error of measurement and intra-class correlation coefficients (ICC) were calculated. Pearson correlations were also calculated to assess the relationships between functional test kinematics and jogging kinematics.

Table 1: Description of the five functional tests used in the study.

Functional test	Test description
Small knee bend (SKB)	In double leg stance, participants were instructed to maximally dorsiflex without lifting their heels.
Single leg SKB (dominant leg)	In single leg stance, participants were instructed to maximally dorsiflex without lifting their heel.
Lunge (dominant leg)	Participants were instructed to step forward a distance of approximately one and a half times the length of their normal gait stride. As they moved into single leg stance they were instructed to maximally dorsiflex without lifting their heel.
Hop lunge (dominant leg)	Starting in double leg stance, participants were instructed to jump forward a distance of approximately 1.0 m and on landing to maximally dorsiflex without lifting their heel.
Step-down	Leading with the non-dominant leg, participants stepped slowly down from a step 20 cm high.

Results: For the five functional tests within-day ICC range was 0.92 to 1.0 and typical error range was 0.5 to 1.3 degrees. There was more between-day variability with ICC range 0.46 to 0.98 and typical error range 1.2 to 3.9 degrees. Pearson correlations for the relationships between functional tests and jogging pelvic, hip and knee kinematics ranged from 0.57 to 0.95.

Conclusions: Based on these preliminary results it appears that lower extremity functional testing has adequate within-day and between-day reliability for most kinematics of interest. There is also some evidence that these tests are predictive of the kinematics that occur during jogging. Both findings support the use of these tests in the clinic to assess lower extremity movement quality and aid clinical decisions.

Participant Information Sheet

AUT
UNIVERSITY
TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

Date Information Sheet Produced

12 December 2010

Project Title

Relationship of hip strength to running mechanics in young athletes.

(Final stage study: The reliability and validity of physiotherapist observational screening of movement during lower extremity functional tests in young athletes.)

Invitation

Thank you for responding to our advertisement. You are invited to participate in the above research project.

What is the purpose of this research?

The purpose of this stage of the study is to examine the reliability and validity of physiotherapy observational assessment of lower limb function in young athletes during common functional tests. It is hoped that this will lead to the development of a screening test that can identify those individuals at risk of lower limb injury. The study is part of a PhD being undertaken by Chris Whatman who is both a staff member and PhD candidate at AUT University. Results will be presented at a physiotherapy conference and published in a journal.

How are people chosen to be asked to be part of this research?

You have responded to advertisements that you have read in physiotherapy newsletters, on websites or received via email. All qualified physiotherapists who volunteer will be able to take part in the study.

What happens in this research?

You will be provided with access to video clips of 23 young athletes performing common lower extremity functional tests and a standardised rating scale. You will be asked to view the video clips and use the scale to rate pelvic and lower limb frontal plane control. You will be asked to complete the ratings of all athletes on two occasions separated by a minimum of two weeks.

What are the discomforts and risks?

There are no discomforts or risks associated with participation.

What are the benefits?

The study will contribute to a larger project that aims to identify individuals at risk of lower limb injury during physical activity and thus aid in the prevention of these injuries.

What compensation is available for injury or negligence?

Compensation is available through the Accident Compensation Corporation within its normal limitations.

How will my privacy be protected?

All participants will be coded for data analysis and presentation of results. Individuals will not be identified in any presentation of the results.

What are the costs of participating in this research?

Ratings of all athletes will take approximately 2 hours on each occasion.

What opportunity do I have to consider this invitation?

If people would like to participate they should contact the researcher as soon as possible after considering this invitation.

How do I agree to participate in this research?

When participants agree to take part in the study they will be required to sign a consent form.

Will I receive feedback on the results of this research?

A summary of findings will be given to those who want it if they have indicated this on the Consent Form.

What do I do if I have concerns about this research?

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisors:

Professor Patria Hume, patria.hume@aut.ac.nz, 9219999 ext. 7306.

Associate Professor Wayne Hing, wayne.hing@aut.ac.nz, 9219999 ext. 7800.

Concerns regarding the conduct of the research should be notified to the Executive Secretary, AUTEC, Madeline Banda, *madeline.banda@aut.ac.nz*, 921 9999 ext. 8044.

Who do I contact for further information about this research?

Researcher Contact Details:

Mr Chris Whatman, chris.whatman@aut.ac.nz, 921 9999 ext. 7307.

Project Supervisor Contact Details:

Professor Patria Hume, patria.hume@aut.ac.nz, 9219999 ext. 7306.

Associate Professor Wayne Hing, wayne.hing@aut.ac.nz, 921 9999 ext. 7800.

Approved by the Auckland University of Technology Ethics Committee on 14 August 2006, AUTEC

Reference number 08/258.

Participant Information Sheet

Date Information Sheet Produced: 22nd October 2008

Project Title How does the strength of my hip affect my running?

An invitation
My name is Kelly Sheerin and I am inviting you to help with a project that looks at how children run. It will involve running on a treadmill and some basic measurements of muscle strength and flexibility. There is also a possibility of doing some simple new exercises as part of your existing Long Term Athlete Development programme. Together, you and your parents should decide whether or not you would like to be involved. You don't have to be involved, it won't affect your role in the programme, and you can stop being involved in the study at any time.

What will happen in this project?
You will come to the University two times on different days, for 1 hour at a time. During each session we will measure your muscle strength, flexibility of your legs and video you running on a treadmill. You will have to run on the treadmill for 5 minutes in total. It will be at your own speed, so you shouldn't be tired at the end of it. You will also do two exercises (a small hop and a small jump). A video recording of these exercises will be shown to some physiotherapists to assess your technique. We will record your height, weight and injury history. In addition, one group will undergo some basic strengthening exercises during their normal LTAD sessions.

What chance do I have to decide whether or not I would like to be involved in the study?
You may take the time you need to talk to your parents and decide whether or not you would like to be involved in the project. You can stop being involved in the project at any point. Your involvement in this study is voluntary and it will not affect what you would normally do as part of the programme.

How do I agree to become involved in this research?
If you and your parents decide that you would like to be involved in the study please ask your parents to contact Kelly Sheerin (contact details below).

Will I receive feedback on the results of this research?
Yes, a report will be provided to your parents with your results.

How will my privacy be protected?
All information related to you will coded in order to ensure that you cannot be identified. The information will remain in locked storage and will only be accessible to the people running the project. No-one at the Millennium Institute of Sport and Health will be able to identify you from any findings that they are given.

What happens if there are concerns about this research?
Any concerns regarding the nature of this project should be notified in the first instance to the Primary Researcher:

Primary Researcher Contact Details: Kelly Sheerin, Institute of Sport and Recreation Research New Zealand, School of Sport and Recreation, AUT University. Email: kelly.sheerin@aut.ac.nz or phone +64 9 917 9999 ext. 7354.

Project Supervisor Contact Details: Associate Professor Patria Hume, Institute of Sport and Recreation Research New Zealand, School of Sport and Recreation, AUT University. Email: patria.hume@aut.ac.nz or phone +64 9 917 9999 ext. 7306.

Concerns regarding the conduct of the research should be notified to the Executive Secretary, AUTEC, Madeline Banda, *madeline.banda@aut.ac.nz*, 921 9999 ext. 8044.

Approved by the Auckland University of Technology Ethics Committee on 10/12/2008, AUTEC Reference number 08/258.

Parent / Guardian Information Sheet

Date Information Sheet Produced: 22nd October 2008

Project Title Relationship of hip strength to running mechanics in young athletes.

An Invitation
Your child has been invited to participate in this research study, which aims to determine the relationship of hip strength to running mechanics in children. Their participation in the study is voluntary and that they may withdraw at any time without any adverse consequences. Withdrawal from the study will not affect your's or their relationship with AUT or Millennium Institute of Sport and Health, or their participation in the Long Term Athlete Development programme.

What is the purpose of this research?
Athletic success in any sport requires years of dedicated training and one requirement of effective training is that athletes remain free of injury. However, over one third of young athletes seek medical attention for injuries that occur during physical activity or sport. Injuries sustained in youth have the potential to disturb growth and lead to chronic disability in adulthood, such as arthritis. The purpose of this research is to determine whether a programme designed to increase hip muscle strength can improve running mechanics and prevent the incidence of injuries in young athletes. The data gathered in this study may be presented and conferences and published in academic journals.

How was my child chosen for this study?
Your child has been selected for this study because they are a participant in the Millennium Institute of Sport - Long Term Athlete Development (LTAD) programme.

What will happen in this research?
Your child will be required to attend two testing sessions at The AUT University Running Mechanics Clinic, 90 Akoranga Drive, Northcote. During each session they will be taken through some basic tests of muscle strength and flexibility and will be videoed while they run on a treadmill for 5 minutes and while they perform two functional tests (a small knee bend and a small jump). The video recordings of the functional tests will be shown to 15 experienced physiotherapists for visual rating. In addition, one group will undergo some basic strengthening exercises during their normal LTAD sessions.

Outline of Testing Sessions

Station 1	Station 2	Station 3	Station 4
Study explanation, questions and informed consent (5 mins) Injury history questionnaire (5 mins) Baseline height and weight measures (5 mins)	Strength measures (10 mins) Flexibility measures (10 mins)	Mark-up with reflective markers (10 mins)	Functional screening tests (5 mins) Gait mechanics (10 mins)
15 mins	20 mins	10 mins	15 mins
Total time	60 mins		

What are the discomforts and risks?

Most children will not be familiar with running on a treadmill, so the opportunity will be provided prior to the first testing session for each child to be familiarised with this. The treadmill run will be at a self-selected speed, so children shouldn't be fatigues at the end of the test.

What are the benefits?
Some of the benefits of participating in this research include finding out whether there are any factors related to how your child runs that could contribute to future injury or discomfort.

What compensation is available for injury or negligence?
In the unlikely event of a physical injury as a result of your child's participation in this study, rehabilitation and compensation for injury by accident may be available from the Accident Compensation Corporation, providing the incident details satisfy the requirements of the law and the Corporation's regulations.

How will my child's privacy be protected?
All information and data related to your will coded in order to ensure that they cannot be identified. The data collected will remain in locked storage and will only be accessible to the principal investigators of the project in accordance with the Privacy Act 1993. No-one at the Millennium Institute of Sport and Health will be able to identify participants from any findings that they are given.

What are the costs of participating in this research?
Your child will be required to attend 2 testing sessions, of one-hour duration each. You will be given a petrol voucher to assist with any travel costs involved.

What opportunity do my child and I have to consider this invitation?
You may take the time you need to consider this invitation to participate in the study. Your child has the opportunity to withdraw from the study at any point. Your child's participation in this study should be voluntary.

How do I and my child agree to participate in this research?
If you and your child decide that you would like to be involved in the study please contact Kelly Sheerin (contact details below).

Will I receive feedback on the results of this research?
Yes, a report will be provided to you and your child with your child's individual results.

What do I do if I have concerns about this research?
Any concerns regarding the nature of this project should be notified in the first instance to the Primary Researcher:

Primary Researcher Contact Details: Kelly Sheerin, Institute of Sport and Recreation Research New Zealand, School of Sport and Recreation, AUT University. Email: kelly.sheerin@aut.ac.nz or phone +64 9 917 9999 ext. 7354.

Project Supervisor Contact Details: Associate Professor Patria Hume, Sport Performance Research Institute New Zealand, School of Sport and Recreation, AUT University. Email: patria.hume@aut.ac.nz or phone +64 9 917 9999 ext. 7306.

Concerns regarding the conduct of the research should be notified to the Executive Secretary, AUTEC, Madeline Banda, *madeline.banda@aut.ac.nz*, 921 9999 ext. 8044.

Approved by the Auckland University of Technology Ethics Committee on 10/12/2008, AUTEC Reference number 08/258.

Participant Information Sheet

Date Information Sheet Produced:

21 June 2007

Project Title

Reliability and validity of physiotherapy screening of lower limb function.

Invitation

Thank you for responding to our advertisement. You are invited to participate in the above research project.

What is the purpose of this research?

The purpose of this study is to examine the reliability and validity of physiotherapy assessment of lower limb function during 3 common clinical tests. It is hoped that this will lead to the development of a screening test that can identify those individuals at risk of lower limb injury. The study is part of a PhD and results will be presented at a physiotherapy conference and published in a journal.

How are people chosen to be asked to be part of this research?

You have responded to notices that have been placed around the University recruiting participants for this research. All volunteers (who are not currently students of the researcher) who meet the inclusion criteria and are not excluded by the exclusion criteria (see below) will be able to take part in the study.

Inclusion criteria:
Healthy status, males and females, 18-35 years old.

Exclusion criteria:
A participant will be excluded from the study if they have:

1. The presence of an injury or pain in the lower back or lower extremity.
2. Any chronic lower limb injury that affects lower limb function e.g. ACL injury.

What happens in this research?

Participants will be asked to come to the Gait Laboratory for approximately 30-45 minutes. Participants will be required to wear shorts and have reflective markers attached to the trunk, pelvis and leg. Participants will then be video taped jogging and performing simple movements similar to a double and single leg squat. This video footage will then be visually assessed by a group of physiotherapists.

What are the discomforts and risks?

There are no discomforts or risks associated with participation other than those associated with normal physical activity.

What are the benefits?

The study will contribute to a larger project that aims to identify individuals at risk of lower limb injury during physical activity and thus aid in the prevention of these injuries. Individual participants may benefit from assessment of their own lower limb movement.

What compensation is available for injury or negligence?

Compensation is available through the Accident Compensation Corporation within its normal limitations.

How will my privacy be protected?

All participants will be coded for data analysis and presentation of results. Individuals will not be identified in any presentation of the results.

What are the costs of participating in this research?

Each participant will be required to make their on way to the Gait Laboratory, Akoranga

Campus, AUT and attend for approximately 30-45 minutes.

What opportunity do I have to consider this invitation?

If people would like to participate they should contact the researcher as soon as possible after considering this invitation.

How do I agree to participate in this research?

When participants attend the Gait Laboratory they will be required to sign a consent form.

Will I receive feedback on the results of this research?

A summary of findings will be given to those who want it if they have indicated this on the Consent Form.

What do I do if I have concerns about this research?

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisor, Wayne Hing, *wayne.hing@aut.ac.nz*, 9219999 ext.: 7800.

Concerns regarding the conduct of the research should be notified to the Executive Secretary, AUTEC, Madeline Banda, *madeline.banda@aut.ac.nz*, 921 9999 ext. 8044.

Who do I contact for further information about this research?

Researcher Contact Details:

Mr Chris Whatman, chris.whatman@aut.ac.nz, 921 9999 ext. 7307

Project Supervisor Contact Details:

Dr Wayne Hing, wayne.hing@aut.ac.nz, 921 9999 ext. 7800

Approved by the Auckland University of Technology Ethics Committee on 14 August 2006, AUTEC Reference number

06/143

APPENDIX 4: PARTICIPANT CONSENT FORMS

CONSENT TO PARTICIPATION IN RESEARCH

Title of Project: **Relationship of hip strength to running mechanics in young athletes**

          (Final stage study: The reliability and validity of physiotherapist observational screening of movement during lower extremity functional tests in young athletes).

Project Supervisor: **Dr Wayne Hing**

Researcher: **Chris Whatman**

- I have read and understood the information provided about this research project (Information Sheet dated 12 December 2010)

- I have had an opportunity to ask questions and to have them answered.

- I understand that I may withdraw myself or any information that I have provided for this project at any time prior to completion of data collection, without being disadvantaged in any way.

- I agree to take part in this research.

- I wish to receive a copy of the report from the research: tick one: Yes O No O

Participant signature: ....……………………..

Participant name: …………………………………………………………………

Participant Contact Details (if appropriate):

…………………………………………………………………………………..

…………………………………………………………………………………..

…………………………………………………………………………………..

Date:

Approved by the Auckland University of Technology Ethics Committee on 14

August 2008, AUTEC Reference number 08/258

Note: The Participant should retain a copy of this form.

AUT UNIVERSITY

TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

Parent / Guardian Consent Form

Date Information Sheet Produced: 22nd October 2008
Project Title: Relationship of hip strength to running mechanics in young athletes

Primary Researcher: Kelly Sheerin

- I have read and understood the information provided about this research project in this Information Sheet.
- I have had an opportunity to ask questions and to have them answered.
- I understand that I may withdraw my child/children and/or myself or any information that we have provided for this project at any time prior to completion of data collection, without being disadvantaged in any way.
- If my child/children and/or I withdraw, I understand that all relevant information including tapes and transcripts, or parts thereof, will be destroyed.
- My child/children are not suffering from heart disease, high blood pressure, any respiratory condition, any illness or injury that impairs my physical performance, or any infection.
- I permit the researcher to use the videos that are part of this project and/or any photographs from them and any other reproductions or adaptations from them, either complete or in part, alone or in conjunction with any wording and/or drawings solely and exclusively for research or educational purposes.
- I understand that the videos will be used for academic purposes only and will not be published in any form outside of this project without my written permission.
- I understand that any copyright material created by the video is deemed to be owned by the researcher and that I do not own copyright of any of the video.
- I agree to my child/children taking part in this research.
- I wish to receive a copy of the report from the research (please tick one):

Yes ○   No ○

Child's name :………………………………………………………………………….

Parent/Guardian's signature:...……………………

Parent/Guardian's name:...........……………………………………………….

Parent/Guardian's Contact Details (if appropriate):
……………………………………………………………………………………..
……………………………………………………………………………………..
Date: ...………………
Note: The Participant should retain a copy of this form.
Approved by the Auckland University of Technology Ethics Committee on 10/12/2008, AUTEC Reference number 08/258.

Participant Assent Form

Date Information Sheet Produced: 22nd October 2008

Project Title: How does the strength of my hip affect my running?

Primary Researcher: Kelly Sheerin

- I have read and understood the information provided about this project in the Information Sheet.
- I have had a chance to ask questions and have them answered.
- I understand that I may withdraw myself or any information that I have provided for this project at any time prior to completion of the project activities, without being disadvantaged in any way.
- I will allow the researcher to use the videos and photographs that are part of this project exclusively for research or educational purposes
- I understand that the videos will be used for academic purposes only and will not be published in any form outside of this project without my written permission.
- I understand that any copyright material created by the video is deemed to be owned by the researcher and that I do not own copyright of any of the video.
- I agree to take part in this project.
- I wish to receive a copy of the report from the project (please tick one):
  Yes○   No○

Participant's signature: ....………...........……..

Participant's name: ....…………………………………………………

Participant Contact Details (if appropriate):
…………………………………………………………………………………..

Date:

Note: The participant should retain a copy of this form.

Approved by the Auckland University of Technology Ethics Committee on 10/12/2008, AUTEC Reference number 08/258.

CONSENT TO PARTICIPATION IN RESEARCH

Title of Project: **Reliability and validity of physiotherapy screening of lower limb function**

Project Supervisor: **Dr Wayne Hing**

Researcher: **Chris Whatman**

* I have read and understood the information provided about this research project (Information Sheet dated 21 June 2007)

* I have had an opportunity to ask questions and to have them answered.

* I understand that I may withdraw myself or any information that I have provided for this project at any time prior to completion of data collection, without being disadvantaged in any way.

* I agree to take part in this research.

* I wish to receive a copy of the report from the research: tick one: Yes O No O

Participant signature: ....……………………..

Participant name: …………………………………………………………….

Participant Contact Details (if appropriate):

…………………………………………………………………………………..

…………………………………………………………………………………..

…………………………………………………………………………………..

Date:

Approved by the Auckland University of Technology Ethics Committee on 14

August 2006, AUTEC Reference number 06/143

Note: The Participant should retain a copy of this form.

APPENDIX 5: ETHICS APPROVAL

Ethics Application Number 06/143 - Reliability and validity of physiotherapy screening of lower limb function.

M E M O R A N D U M

To:	Wayne Hing
From:	**Charles Grinter** Ethics Coordinator
Date:	14 August 2006
Subject:	Ethics Application Number 06/143 **Reliability and validity of physiotherapy screening of lower limb function.**

Dear Wayne

I am pleased to advise that a subcommittee of the Auckland University of Technology Ethics Committee (AUTEC) has approved your ethics application at their meeting on 7 August 2006. This delegated approval is made in accordance with section 8.1 of AUTEC's *Applying for Ethics Approval: Guidelines and Procedures* and is subject to endorsement at AUTEC's meeting on 11 September 2006.

Your ethics application is approved for a period of three years until 7 August 2009.

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

- A brief annual progress report indicating compliance with the ethical approval given using form EA2, which is available online through *http://www.aut.ac.nz/research/ethics*, including a request for extension of the approval if the project will not be completed by the above expiry date;

- A brief report on the status of the project using form EA3, which is available online through *http://www.aut.ac.nz/research/ethics*. This report is to be submitted either when the approval expires on 7 August 2009 or on completion of the project, whichever comes sooner;

You are reminded that, as applicant, you are responsible for ensuring that any research undertaken under this approval is carried out within the parameters approved for your application. Any change to the research outside the parameters of this approval must be submitted to AUTEC for approval before that change is implemented.

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.

To enable us to provide you with efficient service, we ask that you use the application number and study title in all written and verbal correspondence with us. Should you have any further enquiries regarding this matter, you are welcome to contact me by email at *charles.grinter@aut.ac.nz* or by telephone on 921 9999 at extension 8860.

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

Yours sincerely

Charles Grinter

Ethics Coordinator

On behalf of Madeline Banda, **Executive Secretary, AUTEC**

Cc: Chris Whatman chris.whatman@aut.ac.nz

Ethics Application Number 08/258 - Relationship of hip strength to running mechanics

in young athletes.

M E M O R A N D U M
Auckland University of Technology Ethics Committee (AUTEC)

To: Patria Hume
From: **Madeline Banda** Executive Secretary, AUTEC
Date: 23 December 2010
Subject: Ethics Application Number 08/258 **Relationship of hip strength to running mechanics in young athletes.**

Dear Patria

I am pleased to advise that I have approved a minor amendment to your ethics application allowing an increase in the number of participants being recruited and alteration of the research team from James Croft (primary), Patria Hume, Kelly Sheerin, and Chris Whatman, to Patria Hume (primary), Kelly Sheerin, Chris Whatman, and Wayne Hing. This delegated approval is made in accordance with section 5.3.2 of AUTEC's *Applying for Ethics Approval: Guidelines and Procedures* and is subject to endorsement at AUTEC's meeting on 24 January 2011.

I remind you 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 10 December 2011;

- 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 10 December 2011 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 Charles Grinter, Ethics Coordinator, by email at ethics@aut.ac.nz or by telephone on 921 9999 at extension 8860.

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

Yours sincerely

Madeline Banda
Executive Secretary
Auckland University of Technology Ethics Committee

Cc: Kelly Sheerin, Chris Whatman, Wayne Hing

AUTEC approval memorandum

To: Patria Hume
From: Madeline Banda Executive Secretary, AUTEC
Date: 10 December 2008
Subject: Ethics Application Number 08/258. Relationship of hip strength to running mechanics in young athletes.

Dear Patria,

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 10 November 2008 and that 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 19 January 2009.

Your ethics application is approved for a period of three years until 10 December 2011.

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/about/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 10 December 2011;
- A brief report on the status of the project using form EA3, which is available online through *http://www.aut.ac.nz/about/ethics*. This report is to be submitted either when the approval expires on 10 December 2011 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 Charles Grinter, Ethics Coordinator, by email at charles.grinter@aut.ac.nz or by telephone on 921 9999 at extension 8860.

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

Yours sincerely

Madeline Banda
Executive Secretary
Auckland University of Technology Ethics Committee
Cc: Kelly Sheerin, Chris Whatman, James Croft