Aalborg University

10th Semester Sports Technology – Master Thesis

Deadline 3th of June 2016

Validation of Visual Examination During Stair Descent and Single Leg Squat Following Knee Arthroscopy: A Pilot Study

Author: Jakob Hansen, group 10206[a]

Supervision: Carsten M. Mølgaard[b] and Uwe G. Kersting[c]

[a] Stud. Cand Scient. Sports Technology at Department of Health Science and Technology, Aalborg University, Aalborg, Denmark.

[b] Ph.d. at Department of Occupational and Physiotherapy Aalborg University Hospital, Aalborg, Denmark.

[c] Professor at Center for Sensory-Motor Interaction (SMI), Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

Objective: To validate visual examination of arthroscopy patients during challenging ambulation tasks, such as stair descent and single leg squat.

Method: Three arthroscopy patients and 12 healthy subjects performed a stair descent and single leg squat task. 3D motion capture measurements and 2D video recordings were collected. From the 3D motion capture measurement kinematics for trunk, pelvis, knee and foot were calculated, along with the kinetics of the knee; external knee adduction moment (KAM) and knee flexion/extension power. The 2D videos were visually rated by three physiotherapists and intra-rater and inter-rater reliability was calculated using first order agreement statistics (AC1). The agreement between the 3D motion capture measurements and the visual ratings were calculated, also using AC1.

Results: The combined intra-rater agreement was almost perfect (AC1 = 0.91-0.94). The inter-rater agreement was slight for the combined mean (AC1 = 0.29) and substantial deviations in agreements were found (range 0.96). Pairwise comparison of the raters showed discrepancy in agreements, with the highest combined mean of AC1 = 0.55 (range = 0.69) and the lowest combined mean of AC1 = 0.09 (range = 1.49). The comparison between the visual examination and the 3D motion capture method, resulted in a moderate combined mean agreement (AC1 = 0.50) with a substantial deviation (0.80). Comparing the individual raters to the 3D motion capture measurements resulted in a good agreement as the highest (AC1 = 0.72; range = 0.88) and a slight agreement as the lowest (AC1 = -0.04; range = 1.49).

Conclusion: The findings from this study imply that visual rating has an almost perfect intra-rater reliability, but a slight inter-rater reliability, when combining ratings from stair descent and single leg squat. The agreement between raters and 3D motion capture measurements were moderate and rater specific. The agreement between the KAM measured and the overall visual ratings, resulted in slight to moderate agreements- Therefore, the overall visual ratings does not seem to comply with the medial knee loadings of the subject.

Introduction

Knee osteoarthritis (OA) is a major burden worldwide predominantly among elderly people[7]. However, the onset of knee OA is accelerated if injuries to the meniscus or anterior cruciate ligament (ACL) are sustained[14,29]. Further, these injuries are treated with arthroscopic knee surgery which also have been found to add risk to an earlier onset of knee OA[12,13.23,34,42]. Hence, injuries to either the meniscus or the ACL, regardless if surgery is needed or not, can cause earlier onset of knee OA. Meniscal injuries and the severity of these have been associated with the external knee adduction moment (KAM)[8]. In regards to the ACL, Zabala et al.[45] found that ACL reconstructions alter the KAM, as the ACL reconstructed knee has a decreased KAM and the contralateral knee has an increased KAM compared to healthy subjects.

Following arthroscopy surgery, a physiotherapist (typically) evaluates muscle strength, joint mobility, pain and swelling of the joint using visual examination[24]. To evaluate this the patient is asked to perform functional movements such as step board exercises and/or stair descents, which are movements used for evaluation of both range of motion (RoM) and muscle strength[5,15,21,24,41]. Based on the visual examination rehabilitation exercise programs are prescribed to the patients[3,5,25]. Such visual examination methods are widely used in clinical practice[25], but the reliability and validity of these examinations have been investigated by different researchers with discrepancies between the results[10,18,26,39,43]. These discrepancies might result from using different movement tasks or different visual rating methods. Using more challenging movements would likely cause more evident movement abnormalities, compared to regular gait analysis where the differences might be too small to be clearly determined visually. To ensure a reliable and valid examination a standardization of the movements and evaluation forms have been suggested with different methods proposed[6,9,11,35,44]. However, some of the methods focus on only the knee, the lower body or the upper body[6,9,11,] whereas others examine the whole body[35,44]. As trunk tilt in the frontal plane, lateral displacement of the pelvis, pelvis tilt in the frontal plane, knee malalignment in the frontal plane, knee power and foot pronation all influence KAM[1,19,20,22,27,30,31,32,33,36,40,], the examination method should include all these segments. Whatman et al.[44] developed and reliability tested a visual examination method to evaluate all these segments along with oscillations and an overall rating. Results showed acceptable intra-rater and inter-rater reliability especially among experienced physiotherapist. Additionally, the rating form included both a dichotomous and ordinal rating scales, where the dichotomous rating scale showed the best intra-rater and inter-rater agreement. However, Whatman et al.[44] did only test the reliability of the rating without testing the validity of the rating method.

Therefore, the purpose of this study was to test the validity of the rating form proposed by Whatman et al.[44] for challenging ambulation tasks such as, i.e., a stair descent and a single leg squat for both healthy subjects and knee arthroscopy patients. Kinematics for trunk, pelvis, knee and foot along with the kinetics of the external KAM and knee power will be calculated and compared to physiotherapist ratings of the same movements. The level of agreement for intra-rater and inter-rater reliability, along with the level of agreement between the two methods will be determined.

Method

The present study was conducted in the gait laboratory at Aalborg University, Denmark. Data were collected for stair descent and a single leg squat task. Both tasks were performed within the same test session.

Participants

Three males participated as the patient group and two females and 10 males participated as the control group. Anthropometric data, along with Knee injury and Osteoarthritis Outcome Score (KOOS) were collected for each subject (Table 1). The subjects for the patient group were recruited in collaboration with Aalborg University Hospital. All patients who underwent arthroscopy after an ACL injury or a meniscal tear within 8 weeks prior to testing were contacted and asked if willing to participate. The subjects for the patient group were excluded if they had surgery in both knees, had severe pain in the lower extremities besides from the operated knee or were not able to walk on stairs due to pain or joint stiffness. The subjects for the control group were excluded if they had undergone knee surgery, had OA identified in the lower extremities by a MD or if they had severe pain in the lower extremities. Prior to testing, all subjects signed a written consent in accordance to the ethical guidelines of the institution.

Table 1: Anthropometric data for patient group and control group.

	Patient group (n=3)	Control group (n=12)
Age	**47** (± 24) years	**38** (± 19) years
Height	**185** (± 1) cm	**177** (± 8) cm
Weight	**96** (± 14) kg	**77** (± 10) kg
KOOS Pain	**51.9** (± 14.3) score	**98.4** (± 3.8) score
KOOS Symptom	**45.2** (± 10.9) score	**94.9** (± 4.9) score
KOOS ADL	**59.3** (± 16.1) score	**99.8** (± 0.6) score
KOOS Sport/Rec	**20.0** (± 18.0) score	**97.9** (± 4.0) score
KOOS QOL	**20.8** (± 3.6) score	**96.9** (± 7.3) score

Study protocol

Data collection of each subject took place at the same test session and the subject performed both a stair descent task and single leg squat task. Before testing, a full body (excluding arms and head) marker set was applied to the subjects in accordance to Capozzo et al.[4]. For the upper body markers were placed on the 7th cervical vertebra, suprasternal notch, xiphoideus, 12th thoracic vertebrae, bilaterally on the acromion, anterior superior iliac spine, on the posterior superior iliac spine and on the iliac crest. For the lower body (including both legs) the markers were placed on the trochanter, the lateral femoral condyles, the medial femoral condyles, the lateral malleoli, the medial malleoli, calcaneus and first, second and fifth metatarsal heads. Further, clusters consisting of four markers were attached laterally to both thighs and both lower legs. After applying markers, a static reference trial was collected, whereupon the trochanter markers were removed as those were placed on clothes and therefore considered unreliable during the dynamic trials.

Stair descent

The stair descent task was performed on a custom build staircase consisting of five steps. The first three steps had a width of 1000 mm, a depth of 250 mm and a step height of 180 mm. For the last two steps force platforms were integrated into the staircase. The force platforms were located in the centre of the last two steps leaving the same step height but a modified depth of 230 mm. Approximately 2 m of track on a plane surface followed the staircase (Figure 1).

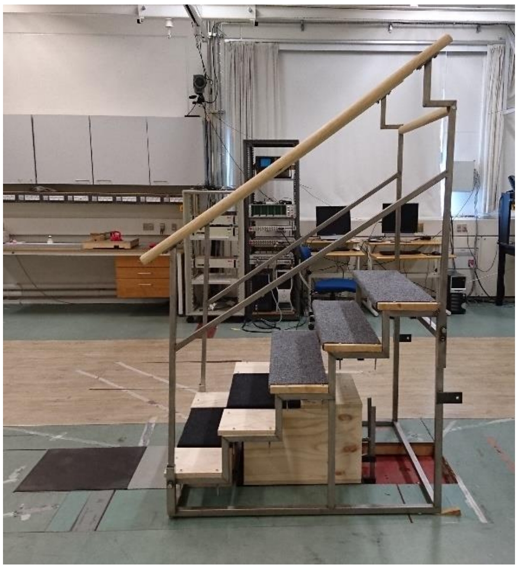

Figure 1: The staircase used in the present study.

Prior to collecting data for the stair descent trials the subjects had time for familiarization to the stairs.

When the subjects felt comfortable walking on the stairs, the pace of the subjects were set by the use of a metronome. The subjects were asked to follow the metronome which was adjusted if necessary. After familiarization five successful trials were collected for both start on the right and left leg. The subjects were asked to walk down the stairs barefooted, with one leg at each step and to follow the pace of the metronome. A trial was discarded if the subjects had double touchdown on a single step, if the pace deviated distinctively from the metronome or if the data was otherwise unacceptable. 2D video cameras were placed to capture the frontal and sagittal view of the subjects at a distance of approximately 2 m at a height of 50 cm.

Single leg squat

Subsequent to the stair descent task, a single leg squat task was performed on a custom made box with the surface 464 x 508 mm and step height of 140 mm. (Figure 2).

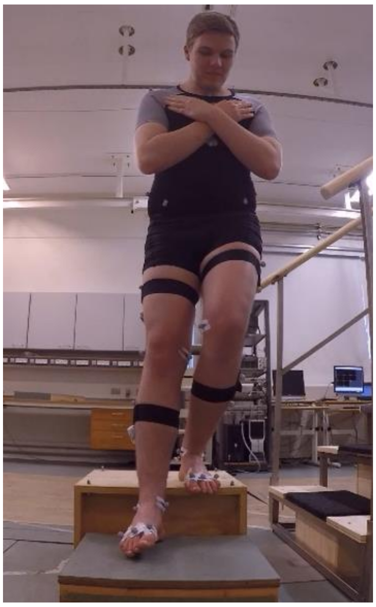

Figure 2: The single leg squat task.

The subjects were asked to do a single leg squat from the box with the non-weight bearing leg reaching for the surface in front of the box. Before data collection the subjects had familiarization trials until comfortable with the movement. The single leg squat was performed barefooted, with the arms crossed in front of the chest and with the instructions to do the movement slow and controlled all the way through. The heel should touch the ground, without transferring weight to the foot. After familiarization five successful trials on the left and right leg were collected. Trials were discarded if the heel did not touch the ground, if balance was lost, if the arms were not crossed in front of the chest or if the data was otherwise unacceptable.

2D video cameras were placed at a distance of approximately 1m and at a height of 50 cm.

Data acquisition

Motion capture data were collected using a setup consisting of eight cameras (Qualisys AB, Gothenburg, Sweden). All cameras were of the Oqus 300 series and had a sample rate of 250 Hz.

GoPro Hero 4 video cameras (GoPro A/S, California, USA) were used to record the 2D videos to be rated by the physiotherapists. The video cameras recorded with 60 Hz and were located in respect to capturing the coronal and sagittal plane of the subject when descending stairs and performing single leg squat. Ground Reaction Forces (GRF) were obtained using three AMTI force platforms (Advanced Mechanical Technology Inc., Watertown, USA). The force platforms were an OR6-5-2000 model and two OR6-7-1000. All GRF data were collected at a sample rate of 1000 Hz with a gain of 2000 and an analog low-pass filter at 1050 Hz.

Data processing

The kinematic and kinetic data collected were processed using C-motion Visual 3D modelling software (Visual 3D, version 5.02.26, C-motion Inc., Germantown, Maryland, USA). A 15 Hz digital low-pass filter was applied to both kinematic and force data and start and end of the movement was defined. For the stair descent trials start and end were defined as above/below 20 N in vertical direction for the platforms on the second last and last step on the stairs. For the single leg squat, start was first frame when all weight was transferred to one leg and end was last frame with weight on one leg. However, this was visually determined. Trunk, pelvis, knee and foot angles, knee power and knee moments were calculated and exported as ASCII files for further processing in MATLAB (v. R2015b, MathWorks, Inc, Nattick, MA, United States). Using MATLAB peak angles, RoM, peak joint power, power integral, peak moment, moment impulse and moment impulse per second were calculated for each of the last two steps on the stairs and for the single leg squat. Further, fast Fourier transform waveform analyses were applied for the trunk abduction/adduction, pelvis tilt, knee varus/valgus and foot pronation/supination movements.

Visual rating

To rate the subjects based on the 2D videos, three physiotherapists with a mean of 7 years of experience (± 5) participated as raters in the study. The subjects of the patient group were evaluated on the leg that underwent surgery, whereas the subjects of the control group were randomly assigned to left or right leg. The physiotherapists rated the videos without any supervision, but with the instructions that they should use the rating form proposed by Whatman et al.[44] (Table 2), could watch the videos as many times as necessary, were not allowed to alter the speed of the videos or pause and were not allowed to use any video software to evaluate angles, etc. The rating consisted of segmental rating and an overall rating. Both rating methods were noted as either normal or abnormal movement.

Table 2: The rating form used for visual ratings.

Segment	*Description*	*Rating 1*	*Rating 2*	*Rating 3*	*Rating 4*	*Rating 5*
Trunk	Moves out of neutral in frontal or transverse plane.	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]
Pelvis 1	Moves out of neutral in frontal or transverse plane.	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]
Pelvis 2	Moves away from the midline.	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]
Knee	Patella moves out of line with 2nd toe.	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]
Foot	Moves into excessive pronation.	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]
Oscillations	Observable oscillation (movement to and from neutral).	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]	[N] [Y; 1, 2, 3]
Overall movement quality	Acceptable movement pattern.	0	0	0	0	0
	Minor movement dysfunction.	1	1	1	1	1
	Moderate movement dysfunction.	2	2	2	2	2
	Marked movement dysfunction.	3	3	3	3	3

Evaluating ratings and measurements

The physiotherapists rated five videos for each of the three tasks per subject; stair descent starting on the left leg, stair descent starting on the right leg and the single leg squat. In order to get an overall evaluation of the three tasks per subject, the ratings of the five trials were merged into one. This was done by defining the category with the most ratings (at least 3/5) as the overall rating.

For intra-rater reliability, two of the three physiotherapists rated seven randomly chosen subjects twice with approximately one week between the ratings. The overall evaluation of both segmental and overall movement ratings were compared for the two rating sessions. For inter-rater reliability between the three physiotherapists, the overall evaluation of both segmental and overall movement ratings where compared for 15 subjects. Inter-rater reliability was evaluated between all of the three raters and additionally compared pairwise.

To compare the visual ratings of the physiotherapists to the 3D motion capture measurements, the overall evaluations of the physiotherapists were merged once again. The category rated by two or more physiotherapists were chosen as the category representing the rating of the physiotherapists. These ratings where compared to the 3D motion capture measurements at second last and last step on the staircase and for the single leg squat (Table 3).

Table 3: Comparison between visual rating and 3D motion capture parameters

Segment	Visual rating description	3D motion capture measurements
Trunk	Moves out of neutral in frontal or transverse plane.	Trunk abduction peak. Trunk abduction RoM.
Pelvis	Moves out of neutral in frontal or transverse plane.	Pelvis tilt peak. Pelvis tilt RoM.
Pelvis 2	Moves away from the midline.	Pelvis lateral displacement RoM
Knee	Patella moves out of line with 2nd toe.	Knee varus/valgus peak. Knee varus/valgus RoM.
Foot	Moves into excessive pronation.	Foot pronation (opposite) peak. Foot pronation (opposite) RoM.
Oscillation	Observable oscillation (movement to and from neutral).	Trunk mean frequency. Pelvis mean frequency. Knee mean frequency. Foot mean frequency.
Overall rating	Acceptable movement pattern. Minor movement dysfunction. Moderate movement dysfunction. Marked movement dysfunction.	KAM peak. KAM impulse. KAM impulse per second. Knee power peak. Knee power integral.

Statistical analysis

To compare the intra-rater and inter-rater reliability the first order agreement (AC1) was calculated in accordance to Blood & Spratt[2] who presented four equations (Equation 1-4) to calculate AC1:

$$\pi_q = \frac{1}{n}\sum_{i=1}^{n}\frac{r_{iq}}{r} \qquad (Equation\ 1)$$

where:

- π_q is the probability that a rater classifies an object into category q.
- n is the number of objects rated in total.
- i goes from 1 to n.
- r_{iq} is the number of raters to classify the i'th object into the q'th category.
- r is the total number of raters.

$$p_{e\gamma} = \frac{1}{Q-1}\sum_{q=1}^{Q}\pi_q(1-\pi_q) \qquad (Equation\ 2)$$

where:

- $p_{e\gamma}$ is the the chance-agreement probability.
- Q is the number of categories in the rating form.

$$p_a = \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{q=1}^{Q}\frac{r_{iq}(r_{iq}-1)}{r(r-1)}\right\} \qquad (Equation\ 3)$$

where:

- p_a is the overall agreement probability.

$$AC1 = \frac{p_a - p_{e\gamma}}{1-p_{e\gamma}} \qquad (Equation\ 4)$$

Thereby, the calculation of the AC1 was corrected for the chance that the physiotherapists would rate the same just because of the only two outcome possibilities. Therefore, the AC1 also range from -1 to 1, where 1 is complete agreement that cannot be explained by change, -1 is complete disagreement that cannot be explained by change and 0 is incomplete agreement/disagreement which is likely to be caused by chance. The scale used to define the strength of the agreements are in accordance to the one used by Whatman et al.[44]:

- 0 ≤ 0.2 = slight
- >0.2 ≤ 0.4 = fair
- >0.4 ≤ 0.6 = moderate
- >0.6 ≤ 0.8 = good
- >0.8 ≤ 1.0 = almost perfect

In regards to comparing the 3D motion capture data and the physiotherapist ratings the same statistical analysis was used. The 3D motion capture data were converted from continuous data to dichotomous data in order to compare the ratings of the physiotherapists. This conversion was performed by establishing a threshold between normal and abnormal movements. The threshold was set as two times the standard deviation of the mean of the control group. Thereby, everything outside the threshold were compared to the abnormal movement ratings of the physiotherapists and everything inside the threshold were compared to the normal movement ratings of the physiotherapists.

For intra-rater and inter-rater reliability of the physiotherapist ratings, the AC1 was calculated for both the segmental and overall rating. Also, a mean and range were calculated of the tasks performed; stair descent starting on the left leg, stair descent starting on the right leg and single leg squat. Further, a combined mean and range for the three tasks were calculated. For the agreement between the visual rating and the 3D motion capture measurements, the AC1 was calculated for each parameter evaluated. A mean and range for agreements at both second last and last step on the stairs, along with agreements for the single leg squat were calculated. Again, a combined mean and range of the three were calculated.

Results

Regarding the physiotherapists rating of oscillation, the data collection was incomplete and could therefore not be included in the study.

Of the three raters, Rater 1 and Rater 3 did the rating twice on half the subjects (Table 4). Except for Rater 1's rating of the pelvis segment during the single leg squat, all AC1 values range within good to almost perfect. All mean ratings for the two raters were classified as almost perfect, but the range differs between the two within the single leg squat, and hence the combined range also differs (Rater 1 = 0.65 and Rater 2 = 0.26).

Table 4: Intra-rater agreement.

Intra-rater agreement	**AC1 Rater 1**	**AC1 Rater 3**
Stair left		
Trunk	0.78	0.78
Pelvis	1.00	0.74
Pelvis 2	1.00	1.00
Knee	1.00	1.00
Foot (opposite)	1.00	1.00
Overall	1.00	1.00
Mean	**0.96**	**0.92**
Range	**0.22**	**0.26**
Stair right		
Trunk	1.00	1.00
Pelvis	1.00	0.74
Pelvis 2	1.00	1.00
Knee	1.00	1.00
Foot (opposite)	1.00	1.00
Overall	1.00	1.00
Mean	**1.00**	**0.96**
Range	**0.00**	**0.26**
Single leg squat		
Trunk	0.74	0.74
Pelvis	0.35	1.00
Pelvis 2	1.00	1.00
Knee	0.74	1.00
Foot	0.84	1.00
Overall	1.00	1.00
Mean	**0.78**	**0.96**
Range	**0.65**	**0.26**
Combined Mean	0.91	0.94
Combined Range	0.65	0.26

The overall inter-rater agreement and the pairwise inter-rater agreement between the three raters showed good to almost perfect agreements for the trunk segment in the two stair descent trials and slight to fair agreements for the rests (Table 5). This resulted in fair mean agreements for the three tasks in a fair combined mean (AC1 = 0.29) with a large range (0.96).

The pairwise comparison showed a combined mean agreement between Rater 1 and Rater 2 is 0.09 with (0.23-0.35) with large ranges (0.67-0.79). This also resulted a range of 1.49, between Rater 1 and Rater 3 the mean was 0.55 with a range of 0.69 and between Rater 2 and Rater 3 the mean was 0.27 with a range of 0.92.

Table 5: Inter-rater agreement.

Inter-rater agreement Physiotherapists	**AC1 Overall**	**AC1 Rater 1 vs Rater 2**	**AC1 Rater 1 vs Rater 3**	**AC1 Rater 2 vs Rater 3**
Stair left				
Trunk	0.82	0.72	0.92	0.80
Pelvis	0.13	-0.28	0.34	0.49
Pelvis 2	0.27	-0.03	0.76	-0.03
Knee	0.14	0.12	0.41	-0.06
Foot (opposite)	0.34	0.48	0.48	0.04
Overall	0.03	-0.44	0.34	0.38
Mean	**0.29**	**0.10**	**0.54**	**0.27**
Range	**0.79**	**1.16**	**0.58**	**0.87**
Stair right				
Trunk	0.78	0.76	0.85	0.72
Pelvis	0.25	0.04	0.34	0.47
Pelvis 2	0.27	-0.03	0.76	-0.03
Knee	0.2	-0.03	0.41	0.33
Foot (opposite)	0.54	0.61	0.72	0.27
Overall	0.05	-0.44	0.48	0.23
Mean	**0.35**	**0.15**	**0.59**	**0.33**
Range	**0.73**	**1.20**	**0.50**	**0.75**
Single leg squat				
Trunk	0.43	0.28	0.56	0.47
Pelvis	0.2	0.2	0.23	0.23
Pelvis 2	-0.14	-0.73	0.61	-0.12
Knee	0.13	-0.28	0.73	-0.03
Foot	0.52	0.54	0.41	0.66
Overall	0.21	0.08	0.49	0.07
Mean	**0.23**	**0.01**	**0.51**	**0.21**
Range	**0.67**	**1.27**	**0.50**	**0.78**
Combined Mean	0.29	0.09	0.55	0.27
Combined Range	0.96	1.49	0.69	0.92

The comparison of the raters to the 3D motion capture measurements showed a combined mean AC1 of 0.5 with a range of 0.8 (Table 6). For second last step on the stairs, good to almost perfect AC1 values were shown for trunk abduction peak, trunk abduction RoM, pelvis lateral displacement RoM, knee varus/valgus RoM, foot pronation peak and foot pronation RoM. For the last step on the stairs good to almost perfect AC1 values were shown for trunk abduction peak, trunk abduction RoM, pelvis lateral displacement RoM, foot pronation peak and foot pronation RoM. For the single leg squat pelvis lateral displacement RoM, foot pronation peak and foot pronation RoM showed good to almost perfect agreements.

Comparing each rater to the 3D motion capture measurement showed a mean (range), for the second last step on the stairs of 0.91 (0.24) for Rater 1, -0.09 (1.29) for Rater 2 and 0.42 (0.88) for Rater 3. For the last step on the stairs the mean (range) were 0.84 (0.39) for Rater 1, -0.06 (1.25) for Rater 2 and 0.40 (0.81) for Rater 3. For the single leg squat the mean (range) were 0.40 (0.72) for Rater 1, 0.01 (1.46) for Rater 2 and 0.33 (0.48) for Rater 3. The combined mean (range) were 0.72 (0.88) for Rater 1, -0.04 (1.49) for Rater 2 and 0.38 (0.88) for Rater 3.

Table 6: Agreements for comparison of visual ratings and 3D motion capture measurements.

Inter-rater agreement Physiotherapists vs 3D MOCAP	**AC1 Overall**	**AC1 Rater 1**	**AC1 Rater 2**	**AC1 Rater 3**
Second last step				
Trunk abduction peak	0.93	1.00	0.76	0.85
Trunk abduction RoM	0.85	0.93	0.65	0.76
Pelvis tilt peak	0.12	0.85	-0.37	-0.03
Pelvis tilt RoM	0.27	0.93	-0.20	0.12
Pelvis lateral displacement RoM	0.85	1.00	-0.20	0.76
Knee varus/valgus peak	0.54	0.93	-0.12	0.48
Knee varus/valgus RoM	0.65	1.00	-0.20	0.41
Foot pronation (opposite) peak	0.76	0.76	0.54	0.54
Foot pronation (opposite) RoM	0.76	0.76	0.54	0.54
KAM peak	0.41	0.93	-0.54	0.27
KAM impulse	0.41	0.93	-0.54	0.27
KAM impulse per second	0.41	0.93	-0.54	0.27
Knee power peak	0.41	0.93	-0.54	0.27
Knee power integral	0.48	0.85	-0.44	0.34
Mean	**0.56**	**0.91**	**-0.09**	**0.42**
Range	**0.80**	**0.24**	**1.29**	**0.88**
Last step				
Trunk abduction peak	0.85	0.93	0.65	0.85
Trunk abduction RoM	0.85	0.93	0.65	0.85
Pelvis tilt peak	0.41	0.93	-0.20	0.12
Pelvis tilt RoM	0.41	0.93	-0.20	0.12
Pelvis lateral displacement RoM	0.65	0.85	-0.12	0.72
Knee varus/valgus peak	0.41	0.85	0.13	0.34
Knee varus/valgus RoM	0.54	0.93	-0.20	0.48
Foot pronation (opposite) peak	0.85	0.76	0.54	0.76
Foot pronation (opposite) RoM	0.85	0.54	0.54	0.76
KAM peak	0.27	0.93	-0.54	0.12
KAM impulse	0.20	0.76	-0.59	0.04
KAM impulse per second	0.20	0.76	-0.59	0.04
Knee power peak	0.27	0.93	-0.54	0.12
Knee power integral	0.20	0.76	-0.33	0.28
Mean	**0.50**	**0.84**	**-0.06**	**0.40**
Range	**0.65**	**0.39**	**1.25**	**0.81**
Single leg squat				
Trunk abduction peak	0.41	0.76	-0.03	0.27
Trunk abduction RoM	0.41	0.76	-0.03	0.27
Pelvis tilt peak	0.34	0.54	-0.12	0.20
Pelvis tilt RoM	0.34	0.55	-0.12	0.20
Pelvis lateral displacement RoM	0.61	0.85	-0.73	0.61
Knee varus/valgus peak	0.20	0.12	-0.28	0.12
Knee varus/valgus RoM	0.12	0.27	-0.37	0.27
Foot pronation (opposite) peak	0.76	0.27	0.65	0.54
Foot pronation (opposite) RoM	0.83	0.34	0.72	0.61
KAM peak	0.41	0.12	-0.03	0.27
KAM impulse	0.41	0.34	0.20	0.27
KAM impulse per second	0.41	0.34	0.20	0.27
Knee power peak	0.54	0.27	0.12	0.41
Knee power integral	0.41	0.12	-0.03	0.27
Mean	**0.44**	**0.40**	**0.01**	**0.33**
Range	**0.70**	**0.72**	**1.46**	**0.48**
Combined Mean	0.50	0.72	-0.04	0.38
Combined Range	0.80	0.88	1.49	0.88

Discussion

The aim of this study was to validate a rating form proposed by Whatman et al.[44].

Three physiotherapists rated 2D videos of 15 subjects performing stair descent and single leg squat. The subjects' movement were rated as either normal or abnormal based on a segmental and overall rating. Both intra-rater and inter-rater reliability was calculated, along with the agreement between visual rating and 3D motion capture measurements.

Intra-rater and inter-rater reliability

The intra-rater reliability showed good to almost perfect agreements except for Rater 1's pelvis rating in the single leg squat (AC1 = 0.35). The comparison of the overall rating to the segmental rating showed that the overall rating was fully reproduced for both stair descent and single leg squat (AC1 = 1), whereas small differences (besides Rater 1's pelvis rating at single leg squat) occurred in the segmental rating (AC1 = 0.74-1). This suggests that the raters are able to reproduce the same results, when evaluating the same movement, especially when evaluating the overall movement. Comparing these results to the findings of Whatman et al.[44] the agreement in the present study is similar or higher, despite present study evaluated stair descent and single leg squat, whereas Whatman et al.[44] investigated small knee bend, single leg small knee bend, lunge and hop lunge. Further, comparing to Chmielewski et al.[6] and Ekegren et al.[11], who investigated single leg squat and drop jump, respectively, the results of the present study show higher intra-rater reliability.

The inter-rater reliability between the three raters showed mixed results. The best agreements were found for the trunk segment during stair descent (AC1 = 0.78-0.82), whereas a slight disagreement were found for the lateral displacement of the pelvis during the single leg squat (-0.14). In general, the highest agreements were found with the segmental rating compared to the overall rating. These results suggest that the discrepancy between the raters are quite large. This can also be seen by the pairwise comparison of the raters, where the best agreement was found between Rater 1 and Rater 3 (mean AC1 = 0.55), suggesting that the ratings of Rater 2 varies the most compared to the others.

Comparing these results to the findings of Whatman et al.[44] a similar or higher agreement in the present study regarding trunk and foot ratings was shown, while the remaining segmental ratings and the overall rating were lower. In general, the inter-rater agreements from the present study was lower than what was found by Ekegren et al.[11] and similar to the findings of Chmielewski et al.[6].

Only a few raters were included in the present study, in regards to both intra-rater (two) and inter-rater (three) reliability. With respect to study design, e.g., Chmielewski et al.[6] and Ekegren et al.[11] this is comparable, as they both used three raters for intra-rater and inter-rater reliability. However, these studies along with the present study lacks power in this regard, when compared to Whatman et al.[44], who used 44 physiotherapists as raters grouped in three different categories of experience. Nonetheless, Whatman et al.[44] discusses the limitations of rating only a homogenous healthy group of subjects. The present study compared a more heterogeneous group as the healthy control group varied substantially in regards to age, height and weight. Further, arthroscopy patients (ACL reconstruction or menisectomy) were included, which might have resulted in a broader range of movements, as knee flexion moment, joint flexibility, tibial internal rotation and knee joint kinematics have been found to change following surgery[16,17,29,38].

Comparing ratings and measurements

The results from comparing the visual ratings to the 3D motion capture measurements across the three points of interest (second last step and last step on the stairs, along with the single leg squat) showed a combined mean agreement of 0.50 with a range of 0.80. This implies that the raters agree moderately with the 3D motion capture system, however with a substantial range in agreements depending on the parameter evaluated.

For the second last step on the stairs, a combined mean agreement of 0.56 with a range of 0.80 was found. Good and almost perfect agreements were found for the trunk abduction peak and RoM, the pelvis lateral displacement RoM, the knee varus/valgus RoM and for both foot pronation peak and RoM. It should be noted that the agreement of the pelvis tilt peak was slight (AC1 = 0.12), suggesting that agreement depend on the parameter evaluated. For the last step on the stairs, a moderate mean agreement was found (AC1 = 0.50), but with a large range (0.65). Good and almost perfect agreements were found for the trunk abducktion peak and RoM, the pelvis lateral displacement RoM and for the foot pronation peak and RoM. As for the second last step, large differences occurred depending on the parameter evaluated. For both KAM impulse, KAM impulse per second and knee power integral slight agreements were found (AC1 = 0.2). Thereby, the agreements for the second last and last step have the same tendencies in terms of parameters evaluated. In regards to the single leg squat, a moderate mean agreement was found (AC1 = 0.44; range 0.70). Good and almost perfect agreements were found for the pelvis lateral displacement and the foot pronation peak and RoM. The rest of the measurements ranges from slight to moderate (AC1 = 0.12-0.54). Thereby, less measurements had an agreement of good or almost perfect, compared to the second last and last step on the stairs, but the mean agreement was similar.

Further, the results only show slight to moderate agreement when comparing the knee kinetics with the overall visual rating. This implies that the overall rating does not capture the full loading of the knee joint in neither stair descent or single leg squat. Results from this study implies that the best agreements with 3D motion capture measurements are found for the trunk abduction, the pelvis lateral displacement, the knee varus/valgus and the foot pronation. Abnormal movements of these segments have been found to increase KAM[1,19,20,27,30,32,33,36,40] and therefore a correct visual rating of these can indirectly be used to evaluate KAM of a subject.

Comparing each rater to the 3D motion capture measurements, shows that Rater 1 has an almost perfect mean agreement for the second last and last step for the stairs (AC1 = 0.91 and 0.84), with smaller ranges (0.24 and 0.39). This suggests, that Rater 1 agrees with the 3D motion capture system for the stair descent trials. However, the mean agreement is not as good for the single leg squat (AC1 = 0.40; range = 0.72). In general, Rater 2 and Rater 3 agrees less with the 3D motion capture measurements, resulting in a lower mean agreement and a larger range for both second last step, last step during stair descent and the single leg squat.

In the present study a method to compare visual ratings to 3D motion capture measurements has been presented. To the knowledge of the author, only a few other studies investigated similar issues. A study by Krosshaug et al.[26] compared visual examination to 3D motion capture measurements of running and side step and found slight accuracy for the visual rating with both random and systematic errors. Further, Harris-Hayes et al.[18], Stensrud et al.[39] and Whatman et al.[43] all compared visual examination to 2D video analysis in different movements (single leg squat, drop jump, single leg drop jump, small knee bend and single leg small knee bend). Analysing 2D videos consisted in all three cases of using different commercial video drawing software. Using the 2D video analysis for comparison, all three studies found that visual rating was a suitable method for screening purposes. However, a comprehensive study by Eastlack et al.[10], who investigated the rater agreement for a video based approach including 54 raters found an inter-rater reliability of low to moderate.

Limitations

The variability of agreements between the two methods might be influenced by the equipment used and the difference occurring from rating videos compared to patients in a clinic. For example, the 3D

motion capture technology is a valuable tool to evaluate movement, but is not without errors[39]. Therefore, the classification of subject parameters could be wrong if close to the threshold values used when converting the data to dichotomous scaling. Further, unstructured feedback was collected from the raters in regards to the differences when rating videos. In general, the physiotherapists are limited to the two angles captured on video, where they are more moveable during clinical practice. This, of course, influenced the rating. For example, when during the single leg squat on one leg, the sagittal plane camera would capture the lateral side when the subject was measured on the right leg, but the medial side when the subject was measured on the left leg, hence giving different conditions between the subjects. Also the pronation angle of the feet can be difficult to determine during stair descent due to distance and size of the object. Especially one rater draw attention to this, as the monitor he used when rating the videos was small. This was also an issue when rating the pelvis, as the subjects was wearing black shorts, making it difficult to examine the rotation of the segment. Further, the visual rating form combines trunk abduction and trunk rotation, along with pelvis tilt and pelvis rotation. Thereby, when comparing to the 3D motion capture measurements that only evaluated trunk abduction and pelvis tilt, there might be a discrepancy.
Another issue is the part of movement rated during the visual examinations. In example, the stair descent trial consists of five steps and therefore it's likely that the raters focused at different steps. A solution to this, could have been to specify that the rating should occur on the last two steps, as the 3D motion capture measurements are. However, this was not chosen as the opportunity to evaluate the whole movement was thought to represent clinical practice better. In regards to this, it should also be noted that the raters did not only evaluate segments and overall rating in terms of the peak and RoM movement, but also by the variability/oscillation of the movement, however, without being able to note it in the form. This feedback suggests that the original plan of comparing the amount of oscillation to the mean frequency of the segment is missing.
It also has to be noted that using the video rating, the physiotherapists were not completely blinded to who underwent surgery as the scars from this were identifiable when the subjects were close to the cameras. Also it should be noted, that not all raters had experience with this structured way of analysing movements and no training sessions in regards to the rating was performed. Further, all three of the raters did the ratings at home and in their spare time, making monitoring of each rater difficult. Therefore, a document with instructions was handed out to all raters, in order to standardize the rating, but differences is still likely to have occurred.
The 3D motion capture measurements were split in normal and abnormal movement based on the control group. The control group consisted of 12 subjects on which normal data was assumed, but most likely this is too few to represent the normal movements occurring during stair descent and single leg squat. Therefore, it would strengthen the present study, if normal data was found and tested by others before application. However, to the knowledge of the author, this data does not exist for the parameters investigated during stair descent and single leg squat.
The current rating method only evaluates the agreement using a dichotomous rating scale. Whatman et al.[44] found better agreement using this rating scale compared to an ordinal scale ranging from 0 to 3. However, the ordinal scale could still be useful as the more detailed rating could be of interest. Using the dichotomous rating scale and merging the ratings from five trials might not give a fair representation of the subject's ability to perform the movement. In example, if the subject does three acceptable movements, but collapses severely during the other two, using the current method the subject would be rated as able to perform the movement. However, in clinical practice this might not be like that. But by including a graduation of severity, the overall rating of the movement might be

shifted from acceptable to moderate dysfunction of the movement.

Conclusion

The findings of the present study showed that, using the rating form proposed by Whatman et al.[44], good to almost perfect intra-rater agreements can be found. Further, agreements were rather consistent across segments.

In regards to inter-rater reliability, the results were inconsistent. In general, trunk abduction, pelvis lateral displacement and foot pronation ratings were the highest both for stair descent and single leg squat, and might be useful measurements in clinical practice. Also, knee varus/valgus might be useful, but only when evaluating stair descent. Most ratings, however, differed between the raters in the present study and therefore low inter-rater agreements were found.

The comparison of 3D motion capture measurements and visual ratings showed discrepancies between the two methods. Agreements were measurement/segment dependent, but in general the visual ratings are varying from the 3D motion capture measurements, suggesting that visual ratings are interpreted with caution. Also, it is noted that differences in agreement occurred between the individual rater and the 3D motion capture measurement. Especially Rater 1 showed better agreement with the 3D motion capture measurements, compared to Rater 2 and Rater 3.

The link between the overall visual rating and the kinetic measurements only rated between slight and moderate. Therefore, an overall movement rating approach are not suggested. Instead, indirect KAM influences from segmental movement ratings are preferable.

References

1. Barrios J. A., Davis I. S., Higginson J. S. and Royer T. D. Lower Extremity Walking Mechanics of Young Individuals with Asymptomatic Varus Knee Alignment. Journal of Orthopaedic Research 2009. 27:1414-1419.
2. Blood E. & Spratt K. F. Disagreement on Agreement: Two Alternative Agreement Coefficients. SAS Global Forum 2007. 186:1-12.
3. Brindle T., Nyland J. and Johnson D. L. The Meniscus: Review of Basic Principles With Application to Surgery and Rehabilitation. Journal of Athletic Training 2001. 36:2:160-169.
4. Capozzo A., Catani F., Croce U. D. and Leardini A. Position and Orientation in space of bones during movement: anatomical frame definition and determination. Clinical Biomechanics 1995. 10:4:171-178.
5. Cavanaugh J. T. & Killan S. E. Rehabilitation following meniscal repair. Current Reviews in Musculoskeletal Medicine 2012. 5:46-58.
6. Chmielweski T. L., Hodges M. J., Horodyski M., Bishop M. D., Conrad B. P. and Tillman S. M. 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 2007: 37:3:122-129.
7. Cross M., Smith E., Hoy D., Nolte S., Ackerman I., Fransen M., Bridgett L., Williams S., Guillemin F., Hill C. L., Laslett L. L., Jones G., Cicuttini F., Osborne R., Vos T., Buchbinder R., Woolf A. and March L. The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. Clinical and epidemiological research. Ann Rheum Dis 2014. 73:1323-1330
8. Davies-Tuck M. L., Wluka A. E., Teichtahl A. J., Martel-Pelletier J., Pelletier J. P., Jones G., Ding C., Davis S. R. and Cicuttini F. M. Association between meniscal tears and the peak external knee adduction moment and foot rotation during level walking in postmenopausal women without knee osteoarthritis: a cross-sectional study. Arthritis Research & Therapy 2008. 10:3.
9. DiMattia M. A., Livengood A. L., Uhl T. L., Mattacola C. G. and Malone T. R. What Are the Validity of the Single-Leg-Squat Test and Its Relationship to Hip-Abduction Strength. J Sport Rehabil 2005. 14:108-123.

10. Eastlack M. E., Arvidson J., Snyder-Mackler L., Danoff J. V. and McGarvey C. L. Interrater Reliability of Videotaped Observational Gait-Analysis Assissments. Physical Therapy 1991. 71:6:465-472.
11. Ekegren C. L., Miller W. C., Celebrini R. G., Eng J. J. and MacIntyre D. L. Reliability and Validity of Observational Risk Screening in Evaluating Dynamic Knee Valgus. J Orthop Sports Phys Ther 2009. 39:9:665-674.
12. Englund M., Roos E. M. and Lohmander L. S. Impact of Type of Meniscal Tear on Radiographic and Symptomatic Knee Osteoarthritis. American College of Rheumatology 2003 48:8:2178-2187.
13. Englund M. & Lohmander L. S. Patellofemoral osteoarthritis coexistent with tibiofemoral osteoarthritis in a meniscectomy population. Annals of the Rheumatic Diseases 2005. 64:1721-1726.
14. Englund M., Roemer F. W., Hayashi D., Crema M. D. and Guermazi A. Meniscus pathology, osteoarthritis and the treatment controversy. Nature Reviews Rheumatology 2012. 8:412-419.
15. Ericsson Y. B., Dahlberg L. E. and Roos E. M. Effects of functional exercise training on performance and muscle strength after meniscectomy: a randomized trial. Scandinavian Journal of Medicine & Science in Sports 2009. 19:156-165.
16. Georgoulis A. D., Papadonikolakis A., Papageorgiou C. D., Mitsou A. and Stergiou N. Three-Dimensional Tibiofemoral Kinematics of the Anterior Cruciate Ligament-Deficient and Reconstructed Knee during Walking. The American Journal of Sports Medicine 2003. 31:1:75-79.
17. Hall M., Wrigley T. V., Metcalf B. R., Hinman R. S., Cicuttini F. M., Dempsey A. R., Mills P. M., Lloyd D. G. and Bennell K. L. Mechanisms underpinning the peak knee flexion moment increase over 2-years following arthroscopic partial meniscectomy. Clinical Biomechanics 2015. 30:1060-1065.
18. Harris-Hayes M., Steger-May K., Koh C., Royer N. K., Graci V. and Salsich G. B. Classification of Lower Extremity Movement Patterns Based on Visual Assessment: Reliability and Correlation With 2-Dimensional Video Analysis. Journal of Athletic Training 2014. 49:3:304-310.
19. Hunt M. A., Simic M., Hinman R. S., Bennell K. L. and Wrigley T. V. Feasibility of gait retraining strategy for reducing knee joint loading: Increased trunk lead guided by real-time biofeedback. Journal of Biomechanics 2011. 44:943.947.
20. Jackson B.D., Wluka A. E., Teichtahl A.J., Morris M.E. and Cicuttini F.M. Reviewing knee osteoarthritis – a biomechanical perspective. J Sci Med Sport 2004. 7:3:347-357.
21. Jeong H. J., Lee S. H. and Ko C. S. Meniscectomy. Knee Surgery & Related Research 2012. 24:3:129:136.
22. Jones R. K., Zhang M., Laxton P., Findlow A. H. and Liu A. The biomechanical effects of a new design of lateral wedge insole on the knee and ankle during walking. Human Movement Science 2013. 32:596-604.
23. Jørgensen U., Sonne-Holm S., Lauridsen F. and Rosenklint A. Long-Term Follow-Up of Meniscectomy in Athletes. British Editorial Society of Bone and Joint Surgery 1987. 69-B:1:80-83.
24. Klinisk instruks for ambulant fysioterapi til patienter med menisklæsion opereret med reinsertion. Internal clinical instructions at Aalborg University Hospital, Aalborg, Denmark.
25. Krebs D. E., Edelstein J. E. and Fishman S. Reliability of Observational Kinematic Gait Analysis. Physical Therapy 1985. 65:7:1027-1033
26. Krooshaug T., Nakamae A., Boden B., Engebretsen L., Smith G., Slauterbeck J., Hewett T. E. and Bahr R. Gait & Posture 2007. 26:378-385.
27. Levinger P., Menz. H.B., Morrow A. D., Bartlett J. R., Feller J. A. and Bergman N. R. Relationship between foot function and medial knee joint loading in people with medial compartment knee osteoarthritis. Journal of Foot and Ankle Research 2013. 6:33.
28. Lohmander S. L., Englund M., Dahl M. L. and Roos E. M. The Long-term Consequence of Anterior Cruciate Ligament and Meniscus Injuries. The American Journal of Sports Medicine 2007. 35:10:1756-1769.
29. Magyar M. O., Knoll Z. and Kiss R. M. The influence of medial meniscus injury and meniscectomy on the variability of gait parameters. Knee Surg Sports Traumatol Arthrosc 2012. 20:290-297.
30. Moyer R. F., Birmingham T. B., Dombroski C.E., Walsh R. F., Leitch K. M., Jenkyn T. R. and Giffin J. R. Combined Effects of a Valgus Knee Brace and Lateral Wedge Foot Orthotic on the External Knee Adduction Moment in Patients With Varus Gonarthrosis. Physical Medicine and Rehabilitation 2013. 94:103-112.

31. Murray A. M., Thomas A. C., Armstrong C. W., Pietrosimone B. G. and Tevald M. A. The associations between quadriceps muscle strength, power, and knee joint mechanics in knee osteoarthritis: A cross-sectional study. Clinical Biomechanics 2015. 30:1140-1145.
32. Mündermann A., Asay J. L., Mündermann L. and Andriacchi T. P. Implications of increased medio-lateral trunk sway for ambulatory mechanics. Journal of Biomechanics 2008. 41:165-170.
33. Pagani C. H. F., Potthast W. and Brüggemann G-P. The effect of valgus bracing on the adduction moment during gait and running in male subjects with varus alignment. Clinical Biomechanics 2010. 25:70-76.
34. Papalia R., Buono A. D., Osti L., Denaro V. and Maffulli N. Meniscectomy as a risk factor for knee osteoarthritis: a systematic review. British Medical Bulletin 2011. 99:89-106.
35. Piva S. R., Fitzgerald K., Irrgang J. J., Jones S., Hando B. R., Browder D. A. and Childs J. D.Reliability of measures of impairments associated with patellofemoral pain syndrome. BMC Musculoskeletal Disorders 2006. 7:33.
36. Resende R. A., Deluzio K. J., Kirkwood R. N., Hassan E. A. and Fonseca S. T. Increased unilateral foot pronation affects lower limbs and pelvic biomechanics during walking. Gait & Posture 2014. 41:395-401.
37. Schwartz M. H., Trost J. P. and Wervey R. A. Measurement and management of errors in quantitative gait data. Gait & Posture 2003. 20:196-203.
38. Shabani B., Bytyqi D., Lustig S., Cheze L., Bytyqy C. and Neyret P. Gait changes of the ACL-deficient knee 3D kinematic assessment. Knee surg. Sports Traumatol Arthrosc 2015. 23:3259-3265.
39. Stensrud S., Myklebust G., Kristianslund E., Bahr R. and Krosshaug T. Correlation between two-dimensional video analysis and subjective assessment in evaluating knee control among elite female team handball players. Journal of Sports Medicine 2010. 45:589-595.
40. Takacs J. & Hunt M. A. The effect of contralateral pelvic drop and trunk lean on frontal plane knee biomechanics during single limb standing. Journal of Biomechanics 2012. 45:2791-2796.
41. Ugalde V., Brockamn C., Bailowitz Z. and Pollard C. D. Single Leg Squat Test and Its Relationship to Dynamic Knee Valgus and Injury Risk Screening. The American Academy of Physical Medicine and Rehabilitation 2015. 7:229-315.
42. Ward B. D. and Lubowitz J. H. Basic Knee Arthroscopy Part 1: Patient Positioning. Arthroscopy Association of North America 2013. 2:4:e497-e499
43. Whatman C., Hume P. and Hing W. The reliability and validity of physiotherapist visual rating of dynamic pelvis and knee alignment in young athletes. Physical Therapy in Sports 2012a. 14:168-174.
44. Whatman C., Hing W. and Hume P. Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. Physical Therapy in Sport 2012b. 13:87-96.
45. Zabala M. E., Favre J., Scanlan S. F., Donahue J. and Andriacchi T. P. Three-dimensional knee moments of ACL reconstructed and control subjects during gait, stair ascent and stair descent. Journal of Biomechanics 2013. 46: 515-520.