Inter-rater and intra-rater reliability of Kinovea software for measurement of shoulder range of motion
Reham M. Abd Elrahim, Eman A. Embaby, Mohamed F. Ali, Ragia M. Kamel

Background
Goniometry is a tool used frequently for measuring and documenting range of motion (ROM) during a physical therapy examination. With modern innovations in technology, new methods other than the universal goniometer have been applied. Kinovea software is a recent video-based method that uses a virtual goniometer to obtain values for the ROM of joints. However, the software’s reliability in measuring shoulder joint ROM has not been studied.

Purpose
This study was conducted to investigate the inter-rater and intrarater reliability of Kinovea software for measuring shoulder joint ROM in healthy individuals.

Materials and methods
Shoulder joint ROM was measured in 52 healthy individuals (mean±SD age was 26.7±4.2 years) using Kinovea photographic measurements by three trained raters. Intrarater reliability was examined by a single rater within the same day. Shoulder flexion, abduction, and external and internal rotation ROM were measured with the patient in supine position.

Results
The inter-rater reliability ranged from an intraclass correlation coefficient value of 0.95 to 0.98, whereas the intrarater reliability ranged from an intraclass correlation coefficient value of 0.98 to 0.99.

Conclusion
This study showed highly reliable shoulder joint ROM measurements in healthy adults using the Kinovea software.

Keywords:
goniometry, movement, photography, reliability, upper limb

Introduction
Daily activities such as drinking, eating, washing, and brushing depend on motions of the upper extremities [1]. The upper extremities make up a mechanically linked unit, and any immobility of one part of that unit has a negative impact on the entire upper extremity unit [2]. The shoulder has the most mobility in the body, and it allows for orientation of the upper extremities as required [3]. The assessment of shoulder range of motion (ROM) is considered crucial for diagnosing disorders of the shoulder and for evaluating the rehabilitation procedures that may influence shoulder function [4,5].

During clinical settings, an active range of motion (AROM) assessment is used to assess movement and state of function of an impaired upper extremity [6]. A review of the literature revealed articles showing different tools used for measuring the ROM of joints of the human body. These tools ranged from easy, inexpensive tools (such as visual estimation and the universal goniometry) to electromagnetic tracking systems to the most technologically advanced, high-cost tools (3D and 4D motion analysis equipment with sensors fixed on the scapula) [7–12]. Traditionally, ROM assessment is carried out using goniometers or inclinometers [12,13].

Goniometers are easy to apply, low in cost, and do not require data reduction [14]. One major drawback of these traditional ROM assessment methods is that the appropriate application depends on the examining clinician who has the potential for error, especially when dealing with the complexities of the shoulder joint [15]. The reference arm of the goniometer should be fixed by the clinician’s hand while the goniometer’s other arm rotates with the movement of the measured
joint. This is not easy in some situations, such as a shoulder internal rotation (IR) and external rotation (ER) at 90° of abduction in which the reference arm does not refer to a bony landmark [12].

A more detailed and extensive measurement of joint mobility can be obtained with motion capture systems using active or passive markers [16–19]. These systems are suitable for biomechanics or motion analysis laboratories more than therapy clinics because of the systems’ large cost and space requirements that often hinder their usefulness in a clinical setting [15,20]. An ideal measuring system is one that is inexpensive and can be used easily without the need for sensors attached to the body [21].

Kinovea is a free software used for the analysis, comparison, and evaluation of sports and training. It is also suitable for physical education teachers and coaches [22]. Kinovea has many advantages; it is easy to use and does not require physical sensors during the analysis [21]. In addition, the software can be used as a measurement tool for motion analysis. Kinovea was used in previous studies to measure the kicking actions of Taekwondo athletes [23], to measure the position, velocity, and acceleration of the lower limbs in healthy participants [21], and to measure the foot strike angles of novice runners [24].

Recently, the validity and reliability of the Kinovea software for measuring the flight time and height of vertical jumping were studied by Balsalobre-Fernández et al. [25]. In addition, the reliability of Kinovea software in measuring the normal resting facial distances and maximal excursions of the selected markers during facial movements was investigated by Baude et al. [26]. Furthermore, Moral-Muñoz et al. [27] stated that Kinovea software is a highly reliable tool that offers an objective method for evaluating hamstring flexibility. However, the reliability of Kinovea software for measuring shoulder joint ROM has never been studied. Thus, the purpose of this study was to assess the inter-rater and intrarater reliability of Kinovea software in measuring shoulder flexion, abduction, and ER and IR ranges in healthy individuals.

Materials and methods

Design of the study

A cross-sectional observational study was used to determine the reliability of Kinovea software (version 0.8.15, available for download at: http://www.kinovea.org). Active ROM of four shoulder joint movements was evaluated using the Kinovea software. Shoulder flexion, abduction, and ERs and IRs were evaluated with the patient in supine position by three raters. Intrarater reliability was examined by a single rater within the same day.

Study sample

The sample size calculation was carried out based on the work of Walter et al. [28]. For intraclass correlation coefficient (ICC), sample size estimation was based on an α value of 0.05, a β value of 0.20, and an expected ICC value (intrarater and inter-rater reliability) of 0.90, with the minimum value in the one-sided 95% confidence interval (CI) of 0.80, using two replicates of each measurement and three evaluators. Using these parameters and defining the unit of assessment being a shoulder, the estimated sample size required ~46 shoulders to be assessed. Allowing for losses, 55 healthy individuals were invited to participate in the study.

This study was conducted in Al-Haram Hospital, Giza, Egypt. The population included in this study was a convenient sample. Participants were recruited among volunteers from the surrounding community (staff and students). They were verbally asked to participate in the study. Healthy adults who had a full ROM of the shoulder girdle and shoulder joints and had no previous history of musculoskeletal or neurological problems of the dominant upper extremity were included in the study. Their BMI ranged from 18 to 25 kg/m². We excluded participants with dominant upper limb or scapular pathology or surgery. Of the 55 invited participants, 52 (31 male and 21 female) met the inclusion criteria and were selected for the study. Thus, 52 shoulders were assessed by the three raters, with a total of 156 measures. In this study, participants’ mean age±SD was 26.7±4.2 years, the mean height±SD was 167.8±9 cm, and the mean weight±SD was 68.9±7.2 kg. The dominant side was tested. The right side was the dominant side in 47 participants and the left side in five participants.

Raters

Three raters (A, B, and C) who completed a short training (2 h) on the method participated in the study. All raters were physiotherapists with more than 5 years of clinical experience in orthopedic physical examination and goniometry (rater A had 20 years of experience, rater B had 12 years of experience, and rater C had 8 years of experience).

Procedures of the study

This study was approved by the Faculty of Physical Therapy Ethics Committee, Cairo University, Egypt.
Participants were informed about the purpose and procedures of the study before their written consent was obtained. Randomization was carried out by asking the participants to draw a folded paper that was labeled with the tested movement from a box.

Participants were asked to wear light clothing to allow for better identification of the bony landmarks and to avoid motion restrictions. Before recording any measurement, the tested movements were practiced three times to familiarize the participants with the procedure and the motions being measured.

Before photographing the motion, a pen marker was used to draw cross marks on preselected anatomical landmarks on the tested dominant upper limb (acromion process, coracoid, olecranon process, and the lateral epicondyly and midthoracic line, and midshaft of the humerus). Using these marks, we quantified the following four movements that each participant performed at a maximum (end-range) joint movement at each participant's own pace:

**Flexion-AROM:** a cross mark was placed on the lateral aspect of the center of the humeral head approximately below the acromion process (fulcrum). One cross mark was placed along the midshaft of the humerus aligned with the greater tuberosity and lateral epicondyly of the humerus; one additional cross mark was placed along the midline of the thorax [29]. Flexion-AROM was assessed with the participant in supine position on a standard plinth. The arm was actively elevated in a strict sagittal plane with the thumb pointed up toward the ceiling.

**Abduction-AROM:** a cross mark was placed on the coracoid process (fulcrum). One cross mark was placed along the shaft of the humerus, and an additional cross mark was placed along the midline of the thorax [29]. Abduction-AROM was measured with the participant in supine position, as in flexion-AROM. The arm was actively elevated in the strict coronal plane with the thumb pointed up toward the ceiling.

**ER-AROM:** a cross mark was placed at the olecranon process (fulcrum), and another cross mark was placed at the ulnar styloid process [29]. ER-AROM was tested with the participant in supine position. The tested arm was supported on the table at 90° abduction, the elbow was flexed to 90°, and the wrist was neutral. A towel roll was placed under the humerus to ensure neutral horizontal positioning and to approximate the plane of the scapula, and a weighted bag was used to prevent unwanted scapular movements. Once positioned, the participant was asked to rotate the arm back into ER to their available end-range without any discomfort. The participant was instructed not to lift the lower back during this measurement.

**IR-AROM:** IR-AROM was measured following the same procedure used for ER, except that the participant was instructed to internally rotate the arm while maintaining the 90° abducted position.

**Instruments**

Images of each participant were captured using a digital camera (Nikon Coolpix S3200, effective pixels 16 MB; Nikon Corp., Tokyo, Japan) to capture the sagittal and the frontal plane profile of the dominant shoulder. The camera was placed 1.5 m away from the participants on a tripod at a height of 80 cm. To maintain the same distance between the camera and the participants; the tripod was placed on taped markers on the floor. All images were imported into a laptop and analyzed using Kinovea software, which is a free, open-source software created for movement analysis (Kinovea, 0.8.15, http://www.kinovea.org/).

Raters were told to view each image only once so that they would not change their assessment. Each rater took two measurements of every participant, and each rater repeated the assessment 1 week later. On the second day of the analysis, the order of appearance of the images on the computer screen was randomized by an assistant to minimize any learning or order effects. All raters were blinded to the results of each other and to their own consecutive repeated results. The intrater reliability phase of the study was conducted by one single rater (rater C) by measuring joint ROM of the same participants twice within the same day (with a period of 3 h in between).

**Data analysis**

All movements were converted into angles using the virtual goniometer of the software.

**Flexion-AROM**

To measure shoulder flexion angle, a line was drawn from the fulcrum point to bisect the midthorax line (stationary arm). Another line was drawn bisecting the point demarking the midshaft of the humerus (movable arm). The angle of the intersection of the two lines was measured in degrees (Fig. 1).

**Abduction-AROM**

To measure shoulder abduction angle, a line was drawn from the fulcrum point bisecting the point of the midthorax line (stationary arm), and another line
bisected the point demarking the shaft of the humerus (movable arm). The angle of the intersection of the two lines was measured in degrees (Fig. 2).

**ER-AROM**
The ER angle was formed by a line drawn from the fulcrum point through the shaft of the ulna (movable arm) and a line perpendicular to the plinth (stationary arm) (Fig. 3).

**IR-AROM**
The same procedure was used to measure the ER of the shoulder (Fig. 4).

**Statistical analysis**
Descriptive data analyses were calculated as well as the relative reliabilities and 95% CIs were expressed as the ICC3,1 and inter-rater correlation coefficients (ICC2,2). All statistical analyses were carried out using the statistical package (SPSS for Windows, version 20; SPSS Inc., Chicago, Illinois, USA). The reference ranking values for the ICC in the present study were those described by Johnson and Gross [30]: small reliability, 0.25; low reliability, 0.26–0.49; moderate reliability, 0.50–0.69; high reliability, 0.70–0.89; and very high reliability more than 0.90. Absolute reliability was expressed as a standard error of measurement (SEM). Further, the minimal detectable change (MDC), which represents the magnitude of change necessary to exceed the measurement error of two repeated measurements at a specified CI, was calculated for the 95% CI (MDC95).

**Results**
The descriptive statistics (mean±SD) of inter-rater and intrarater measurements using Kinovea software are shown in Tables 1 and 2, respectively. The ICC2,2 for inter-rater reliability of shoulder joint measurements ranged from 0.95 (0.92–0.97) for ER to 0.98 (0.97–0.99) for abduction with the error of measurement between raters ranging from 0.92° (ER) to 1.6° (abduction) (Table 3). The MDC95 values were between 2.6° (ER) and 4.4° (abduction) (Table 3). Overall, measurements taken in abduction and ER showed the potentially lowest and highest meaningful reliability properties, respectively. The intrarater reliability of the measurements made with Kinovea software was verified with ICC values for shoulder measurements ranging between 0.98 (0.98–0.99) for flexion and 0.99 (0.98–0.99) for other movements (Table 3). The SEM was small for shoulder ROM in all directions (Table 3), ranging from 0.61° (ER) to 0.77° (IR). The MDC95 values were between 1.7° (ER) and 2.1° (abduction and IR) (Table 3). Overall, measurements taken in flexion and ER showed the potentially lowest and highest meaningful reliability properties, respectively.
Discussion

This study shows that the video-based measuring method of shoulder ROM using Kinovea software applied on plain video recordings of shoulder joint movements is highly reliable within and between raters. Compared with 2D analysis studies, Kinovea software is free, open-sourced, fast to use, and shows very good advantages in terms of its applicability in clinical settings. In addition, it could be frozen at the appropriate time of joint ROM (http://www.kinovea.org/) [25,26].

Previous literature on Kinovea reported that this method does not require prior experience in video recording and analysis to obtain highly accurate and reliable measurements [21,25]. However, in the current study, raters attended a training session before testing to become familiar with the software and to increase the accuracy of their analyses. This was carried out in accordance with the recommendation of Baude et al. [26], who suggested that formal training session for patients and raters be taken to improve the reliability of Kinovea in the motion analysis of facial muscles other than the frontalis muscles.

Furthermore, in the current study, a cross mark was drawn on the skin to identify the joint fulcrum (center of motion) as a practical method to increase the reliability of the study. This is in agreement with the procedure by Baude et al. [26], who used face paint to draw dots on the face of each participant on preselected anatomical facial markers. In addition, Richardson [24] placed athletic tape markers on participants’ shoes to assist in digitizing the foot strike angles of novice runners, and Damsted et al. [31] used a hip marker to quantify knee and hip angles at the foot strike during running. Furthermore, Moral-Muñoz et al. [27] marked the participants’ skin to measure hip and knee joint angles as measurements of hamstring flexibility.

Kinovea is a video analysis software with a built-in angle selection tool that eliminates the need to print out still photos and uses a digital virtual goniometer to measure the joint angles with a precision of 1° increments [24,32]. Accordingly, it is considered a photography-based goniometry that differs from a clinical goniometry with its drawbacks of fixing the goniometer on the patient to measure the ROM of joints.

The findings of the current study revealed that the ICCs for inter-rater reliability of shoulder joint measurements ranged from 0.95 to 0.98, which exceed those reported by Mullaney et al. [12]. They studied the intrarater and inter-rater reliability of a construction grade digital level compared with the standard universal goniometer for measurements of active assisted shoulder ROM in patients with unilateral shoulder pathology. They found that inter-rater ICCs ranged from 0.62 to 0.79 for the goniometer and from 0.31 to 0.87 for the digital level in the measurement of the noninvolved limb.

With regard to intrarater reliability, measurements with Kinovea software were found to be comparable to the findings of Mullaney et al. [12], who reported that ICC values ranged from 0.91 to 0.99 for both the goniometer and the digital level. In addition, Kolber et al. [13] investigated the intrarater reliability of

Table 1 Descriptive statistics (mean±SD) of inter-rater measurements

<table>
<thead>
<tr>
<th>Joint movement</th>
<th>Rater A</th>
<th>Rater B</th>
<th>Rater C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoulder flexion</td>
<td>163.9±6.8</td>
<td>164.2±6.9</td>
<td>164.4±6.5</td>
</tr>
<tr>
<td>Shoulder abduction</td>
<td>95.7±7.3</td>
<td>96.1±7</td>
<td>96.5±7.1</td>
</tr>
<tr>
<td>Shoulder ER</td>
<td>96.7±7</td>
<td>97.1±7</td>
<td>97±7.2</td>
</tr>
<tr>
<td>Shoulder IR</td>
<td>101.7±6.6</td>
<td>101.7±6.5</td>
<td>102.3±6.7</td>
</tr>
</tbody>
</table>

ER, external rotation; IR, internal rotation.

Table 2 Descriptive statistics (mean±SD) of intrarater measurements

<table>
<thead>
<tr>
<th>Joint movement</th>
<th>First measurement</th>
<th>Second measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoulder flexion</td>
<td>164.4±6.1</td>
<td>164.5±5.9</td>
</tr>
<tr>
<td>Shoulder abduction</td>
<td>95.8±7.2</td>
<td>95.8±7</td>
</tr>
<tr>
<td>Shoulder ER</td>
<td>96±7.4</td>
<td>96.1±7.2</td>
</tr>
<tr>
<td>Shoulder IR</td>
<td>102.3±6.7</td>
<td>102.2±6.7</td>
</tr>
</tbody>
</table>

ER, external rotation; IR, internal rotation.
active shoulder mobility measurements using a digital inclinometer and goniometer. Intrarater analysis suggested excellent reliability for all measurements with both instruments with ICC values ranging from 0.94 to 0.98. In the aforementioned study, the SEM was 2°, whereas in the current study the SEM was less than 1°, and the MDC95 was lower to 2°, indicating superior reliability of the Kinovea software.

In comparison with more advanced measurement tools, Kinovea software showed higher intratester reliability compared with smartphone applications. Mitchell et al. [33] examined the reliability of inclinometry-based and photo-based smartphone applications in the measurement of active shoulder ER as compared with standard goniometry (SG). Intrarater reliability using novice raters ranged from ICC values of 0.79 to 0.81 for the two applications and 0.82 with an SG value. Inter-rater reliability (novice/expert) for the two applications ranged from ICC values of 0.92 to 0.94 with an SG value at 0.91. Both applications were found to be reliable and comparable to SG. A photo-based application potentially offers a superior method of measurement, as visualizing the landmarks may be simplified in this format. In addition, it provides a record of measurement that is in agreement with the advantages of Kinovea software.

Furthermore, in a study by Werner et al. [34] the average ICC for the SG was 0.69; the average ICC for the smartphone clinometer was 0.80. These values are lower than the values recorded by Kinovea in the current study. In addition, abduction had the lowest ICC values across examiners, and ER at 0° and 90° had the highest ICC values, which are in agreement with the current findings. In the aforementioned study, SEM calculations for each measurement modality were used to determine absolute reliability. The average clinometer had the lowest (best) SEM at 3.6°. In general, ER and IR measurements had the highest SEMs, whereas abduction and forward flexion typically had the lowest SEMs. Their findings highlighted that Kinovea yields better reliability compared with smartphones, as it has SEMs less than 2°.

As expected, intrarater reliability values were higher than inter-rater values, which had less variability when the same rater was used. In addition, the values of the SEM and the MDC95 for intrarater reliability were smaller than that for inter-rater reliability, which is in accordance with less measurement variation that is typical when the same rater is used. It is worth mentioning that measurement variation of any measurements or measurement tools, in general, is more likely to be detected than no variation because the nature of reality is such that measurements are rarely perfectly reliable. This is due to multifactorial sources to variation existing within the total measurement system [35].

Measurement variations found in the present study could be attributed to the following factors. First, there is the human factor; the inter-rater variation could be higher than the intrarater variation because the analyses were made by three different raters with different personal characteristics and years of experience. Second, there is the time factor, as the variability within same day measurements could be due to the fact that the measurements were taken on two different occasions. Although these potential sources of error existed, Kinovea software shows excellent reliability and is believed to be a valuable clinical tool.

Kinovea software for shoulder ROM presents several advantages over currently used methods, including visual estimation, classic double-armed goniometer, digital inclinometry, and high-speed cinematography. The free, open-access software is widely available to users, giving it a distinct advantage over other digital inclinometers and more complex measurement tools.

Strengths of the present study are highlighted as follows. First, the three raters were trained with Kinovea software before the study, which could have enhanced consistency in the measurement process.

### Table 3 Indices of inter-rater and intrarater reliability

<table>
<thead>
<tr>
<th>Joint movement</th>
<th>Inter-rater reliability</th>
<th>Intrarater reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC2,2 (95% CI)</td>
<td>SEM (deg.) MDC (deg.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulder flexion</td>
<td>0.97 (0.96–0.98)</td>
<td>1</td>
</tr>
<tr>
<td>Shoulder abduction</td>
<td>0.95 (0.92–0.97)</td>
<td>1.6</td>
</tr>
<tr>
<td>Shoulder ER</td>
<td>0.98 (0.97–0.99)</td>
<td>0.92</td>
</tr>
<tr>
<td>Shoulder IR</td>
<td>0.96 (0.94–0.98)</td>
<td>1.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Joint movement</th>
<th>Inter-rater reliability</th>
<th>Intrarater reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC3,1 (95% CI)</td>
<td>SEM (deg.) MDC (deg.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulder flexion</td>
<td>0.98 (0.98–0.99)</td>
<td>0.74</td>
</tr>
<tr>
<td>Shoulder abduction</td>
<td>0.99 (0.98–0.99)</td>
<td>0.75</td>
</tr>
<tr>
<td>Shoulder ER</td>
<td>0.99 (0.98–0.99)</td>
<td>0.61</td>
</tr>
<tr>
<td>Shoulder IR</td>
<td>0.99 (0.98–0.99)</td>
<td>0.77</td>
</tr>
</tbody>
</table>

CI, confidence interval; ER, external rotation; ICC, intraclass correlation coefficient; IR, internal rotation; MDC, minimal detectable change; SEM, standard error of measurement.
Second, standardization of the procedures also could have minimized random errors. Third, measurements were taken in a random order. Because of the numerous measurements taken with the Kinovea software, we have assumed that it would have been impossible for the raters to remember all of the results and influence the readings. Therefore, we believe this helped in minimizing information bias. Finally, the raters made their measurements on videos previously taken by a research assistant. Only one video was taken, and all of the measurements were performed from that image. On the basis of the work of Ferriero et al. [36], if the picture was taken correctly, it did not significantly influence the inter-rater reliability. This study has the following limitations. The enrolled participants were healthy; consequently, our findings cannot be generalized to other patient population with various musculoskeletal dysfunctions. A major limitation would be assessing participants while dressing light clothing; also, markers movement with skin.

Conclusion
Kinovea software is a reliable tool for measuring shoulder flexion, abduction, and IR and ER-ROM in healthy individuals. Thus, it could be used as a simple alternative to universal goniometry.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

References
22. www.kinovea.org [Cited: 10/01/2013].

