Interrater and Intrarater Reliability of the Active Knee Extension (AKE) Test among Healthy Adults

MOHAMAD SHARIFF A HAMID[1)*], MOHAMED RAZIF MOHAMED ALI[2)], ASHRIL YUSOF[3)]

[1)] *Unit of Sports Medicine, Faculty of Medicine, University of Malaya: 50603 Kuala Lumpur, Malaysia.*
  TEL: +603 79674968, FAX: +603 79677511
[2)] *Department of Orthopaedic Surgery, Faculty of Medicine, University of Malaya*
[3)] *Senior Lecturer, Sports Centre, University of Malaya*

Abstract. [Purpose] The purpose of this study was to determine the reliability of the active knee extension (AKE) test among healthy adults. [Subjects] Fourteen healthy participants (10 men and 4 women) volunteered and gave informed consent. [Methods] Two raters conducted AKE tests independently with the aid of a simple and inexpensive stabilizing apparatus. Each knee was measured twice, and the AKE test was repeated one week later. [Results] The interrater reliability intraclass correlation coefficients ($ICC_{2,1}$) were 0.87 for the dominant knee and 0.81 for the nondominant knee. In addition, the intrarater (test-retest) reliability $ICC_{3,1}$ values range between 0.78–0.97 and 0.75–0.84 for raters 1 and 2 respectively. The percentages of agreement within 10° for AKE measurements were 93% for the dominant knee and 79% for the nondominant knee. [Conclusion] The finding suggests the current AKE test showed excellent interrater and intrarater reliability for assessing hamstring flexibility in healthy adults.
Key words: Hamstring, Flexibility, Range of movement

INTRODUCTION

Hamstring muscle injury is one of the commonest sports injuries among athletes[1, 2]. Based on the literature, risks of hamstring muscle injury include previous injury, strength imbalance, older age, inadequate warm-up, poor quadriceps flexibility and muscle fatigue[1–5]. Another possible cause of hamstring injury is muscle tightness. This factor remains inconclusive, as studies have shown conflicting results[6–10]. Despite this, hamstring flexibility is recommended during routine pre-participation examinations and in deciding athlete readiness to return-to-play following injury[11–13]. Therefore, a simple and reliable method of hamstring flexibility assessment is relevant.

Methods to assess hamstring flexibility include the straight-leg-raising (SLR) test, sit and reach (SR) test, and active knee extension (AKE) test[14–16]. The SLR test specificity has been questioned, as it is also widely used as a neurological test[17]. Further, cinematographic study showed that pelvic rotation may influence the validity of SLR angle measurements[18]. Even though hamstring flexibility assessment is easy using the sit and reach (SR) test, the validity of this test is considered poor[19]. The AKE test is an active test that involves movement at the knee joint, and most considers it safe, as the patient dictates the end point of movement. Further, the AKE test aided by a metal rig and straps to limit pelvic and leg motion (as a stabilizing apparatus) showed a high intrarater correlation coefficient when conducted within 30 minutes[16, 20]. Others though, question the practicality of such method, as the apparatus used by researchers are complicated, rarely available in a clinical setting, and require more than one assessor to conduct the test[21, 22]. Therefore, a simple and reliable method of hamstring flexibility assessment that ensures pelvic and leg stability during measurement is needed.

We designed and made a simple, portable and easy-to-use stabilizing apparatus for use in performance of the AKE test. The apparatus was made using polyvinyl chloride (PVC) hollow tubes. PVC tubes were selected because they are light, easily cut, commonly available in hardware stores, and inexpensive. When used with a universal goniometer, this apparatus allows measurement in the AKE test to be conducted by a single assessor without assistance.

The objective of this study was to determine the interrater and intrarater reliability of the AKE test using a PVC stabilizing apparatus for hamstring flexibility assessment in healthy subjects.

SUBJECTS AND METHODS

The University of Malaya Medical Centre Ethics Committee approved the study (MEC Ref no. 835.11). A convenient sample of 16 healthy participants (10 men and 6 women) with ages ranging from 28 to 39 years was used in this study. Participants were Sports Medicine postgraduates and staff at the Faculty of Medicine, University of Malaya. All participants were free from any orthopaedic or neurological disorders.

We designed a simple, portable apparatus made up of PVC tubes (total cost of USD 6.00) to aid in performance

*To whom correspondence should be addressed.
Email: ayip@um.edu.my

Fig. 1. Flexion of the knee until the thigh touches the horizontal PVC bar

Fig. 2. Knee position during the AKE test

of the AKE test. The apparatus is based on those used by an earlier study[16]. It consisted of a single horizontal bar anchored (removable) by two vertical poles on either side of the plinth. Assessments were conducted at the Sports Medicine Clinic, University of Malaya Medical Centre. To ensure that all participants would be assessed in the standard manner, both assessors attended a half-day workshop on the AKE test using the PVC stabilizing apparatus. The AKE testing procedure was described to all participants, and each provided informed consent. The AKE measurements were taken for both knees. The dominant knee was determined based on the participant's preferred leg when kicking a ball or when performing a single-leg jump, while the other knee was considered the nondominant knee; a previous researcher used a similar definition[23].

A sports physician (SAH) and a physiotherapist (LPC) performed all the AKE tests independently. All assessments were conducted at the Sports Medicine Clinic of the University of Malaya Medical Centre between 9.00 and 11.00 am at room temperature. Participants were advised to avoid any sporting activities on the day of assessment. As previous study demonstrated changes in biomechanical characteristics of collagen and muscle viscoelastic properties, no warming up activities were allowed[24]. Participants were assessed on a plinth in the supine position with both lower extremities extended. Both anterior superior iliac spines were positioned by aligning them with the vertical bars of the apparatus. The lower extremity not being measured was secured to the plinth using a strap across the lower third of the thigh. Each assessor marked the lateral knee joint line with washable ink. From there, two lines were drawn. The first was drawn to the greater trochanter, and another to other was drawn to the apex of the lateral malleolus. The participants were told to flex the hip until the thigh touched the horizontal PVC bar (Fig. 1). While maintaining the contact between the thigh and horizontal PVC bar, the participants were asked to extend the leg as much as possible while keeping their foot relaxed and to hold the position for about 5 seconds. A standard universal goniometer was placed over the previously marked joint axis, and the goniometer arms were aligned along the femur and fibula (Fig. 2).

The AKE measurement was defined as the degree of knee flexion from terminal knee extension. Each knee was measured twice, and the mean angle of the AKE test was used for analysis. All participants attended two testing sessions one week apart to allow for establishment of test-retest reliability of the method. The order in which the rater assessed the participants was randomly assigned in the first session and maintained thereafter.

All data were analyzed using the Statistical Package for the Social Sciences (SPSS) version 20.0 software. Data were described descriptively, and a normality test was performed using the Shapiro-Wilks test. Paired t-tests were performed to compare differences between tests and retest measurements within and between raters.

Two different types of measures of reproducibility were assessed: measures of agreement and measures of reliability. The interrater agreement was quantified by calculating the mean difference between the two raters (rater 1 − rater 2) and the standard deviation (SD) of this difference. Further, the 95% limits of agreement were calculated according to the method of Bland and Altman[25]. These limits represent the range in which 95% of the differences between the two raters fall.

Plots of differences between raters against the corresponding mean of the two raters for each participant were produced to examine homoscedasticity as proposed by Bland-Altman[25]. Further, the frequency of agreement of the raters within 5° and 10° was calculated. Differences exceeding 10° were determined as being unacceptable, as a previous study suggested they are likely to affect decisions on patient management[26].

The reliability was evaluated by computing intraclass correlation coefficients (ICCs), which analyzed the consistency of quantitative measures. An ICC based on a two-way random model was used to establish test-retest reliability, while an ICC based on a two-way mixed model was chosen for intrarater reliability[27]. The ICC value of the standard error of measurement (SEM) was calculated to express the magnitude of disparity between measurements[28]. SEM was calculated using the formula , $SEM = SD_{avg} (\sqrt{1-ICC})$ where SD corresponds to the pooled standard deviation and ICC is the reliability coefficient[29]. A smaller SEM value

Table 1. Descriptive data of participants' physical characteristics

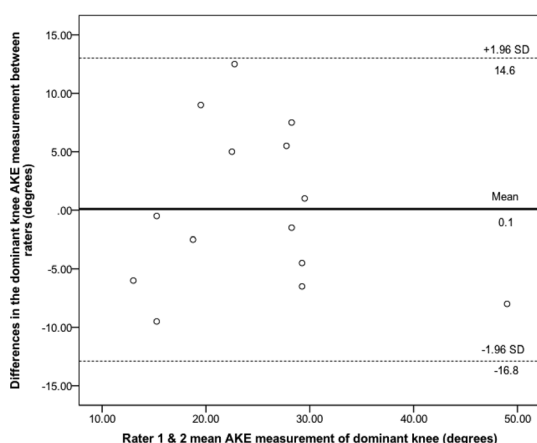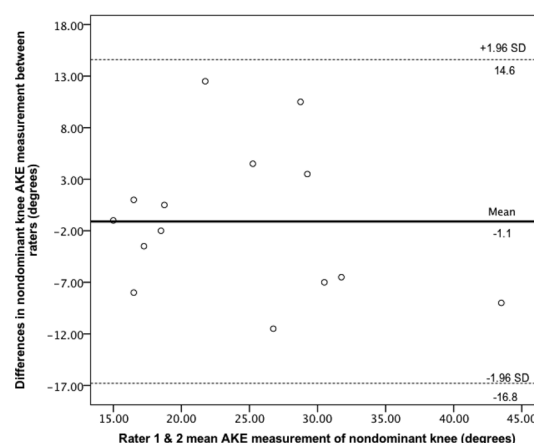| Variables	All participants Mean (SD*)	Range	Male Mean (SD)	Female Mean (SD)
Age (years)	31.5 (2.8)	28–39	32.3 (3.0)	30.5 (2.4)
Body weight (kg)	71.6 (13.5)	54–100	77.7 (12.9)	63.4 (10.3)
Height (cm)	168.4 (6.1)	159–181	170.9 (5.8)	165.0 (4.9)
Body mass index (BMI)	25.1 (3.5)	19.1–31.8	26.5 (3.4)	23.2 (3.0)
AKE† dominant knee (°)	24.9 (9.3)	10.0–45.0	26.6 (10.8)	22.8 (7.2)
AKE nondominant knee (°)	23.7 (8.1)	12.5–39.0	26.1 (9.3)	20.6 (5.3)

*SD, standard deviation
†AKE, active knee extension

Table 2. Mean, SD, and the frequency of agreement within 5 and 10 degrees

AKE* test	Rater 1 Mean (SD*)	Rater 2 Mean (SD)	Rater 1 − 2 Mean (SD)	Upper and lower limit of agreement	Rate of agreement (%)	
					5°	10°
Dominant knee (°)	24.9 (9.3)	24.8 (10.1)	0.1 (6.9)	−13.4 – 13.6	43	93
Nondominant knee (°)	23.7 (8.1)	24.9 (7.2)	−1.1 (7.2)	−15.2 – 13.0	43	79

*AKE, active knee extension
†SD, standard deviation

Fig. 3. Bland-Altman plot of the differences versus means of the dominant knee AKE measurements

Fig. 4. Bland-Altman plot of the differences versus means of the nondominant knee AKE measurements

suggests greater agreement between measurements[30]. The minimum detectable change (MDC) was calculated using the formula MDC = $1.96 \times SEM \times \sqrt{2}$[31].

RESULTS

Fourteen (8 men and 6 women) participants completed the test and retest sessions. Two participants were excluded from the final analysis, as one suffered a hamstring injury and the other was involved in a road traffic accident. A significant difference in body weight between male (mean=77.69, SD=12.86) and female (mean=63.42, SD=10.27) participants was found (t=2.23, p=0.046; Table 1). There was no significant difference in AKE measurements between the dominant and nondominant knees. In addition, no significant difference was observed between test and retest sessions for both dominant (t=0.77, p=0.46) and nondominant (t=−1.01, p=0.33) knees.

A summary of the interrater agreement observed in this study is displayed in Table 2. Both raters had similar AKE measurements, with the limits of agreement being 0.1 ± 12.9 (SD) for the dominant knee and −1.1 ± 15.7 for the nondominant knee. The percentages of agreement within 10° for these measurements were 93% and 79% for the dominant and nondominant knees respectively.

The AKE measurement differences between raters were plotted against the mean value of both raters for both the dominant and nondominant knees (Fig. 3 and 4). Errors of measurement for both knees were independent of the magnitude of the range of measurements (homoscedasticity)

Table 3. Interrater reliability of the AKE test

AKE* measurements	Rater 1, mean (SD†)	Rater 2, mean (SD)	p value	ICC$_{2,1}$‡	95% CI§ ICC	SEM‖	95% CI SEM
Dominant knee (º)	24.9 (9.3)	24.8 (10.1)	0.95	0.87	0.58 – 0.97	3.5	18.0 – 31.7
Nondominant knee (º)	23.7 (8.1)	24.9 (9.5)	0.56	0.81	0.41 – 0.94	3.8	16.9 – 31.7

*AKE, active knee extension test
†SD, standard deviation
‡ICC, intraclass correlation coefficient
§CI, confidence interval
‖SEM, standard error measurement

Table 4. AKE test-retest reliability of raters 1 and 2

AKE* test	Rater	Session 1, mean (SD†)	Session 2, mean (SD)	p value	ICC$_{3,1}$‡	95% CI§ ICC	SEM‖	95% CI SEM
AKE dominant knee (º)	1	24.9 (9.3)	24.0 (6.8)	0.46	0.92	0.76 – 0.97	2.3	19.9 – 29.0
	2	24.8 (10.1)	22.3 (6.2)	0.21	0.78	0.32 – 0.92	3.9	16.0 – 31.2
AKE nondominant knee (º)	1	23.7 (8.1)	25.2 (8.6)	0.33	0.88	0.45 – 0.92	2.9	18.8 – 30.2
	2	24.9 (9.5)	23.1 (5.9)	0.31	0.82	0.46 – 0.94	3.3	17.5 – 30.5

*AKE, active knee extension test
†SD, standard deviation
‡ICC, intraclass correlation coefficient
§CI, confidence interval
‖SEM, standard error measurement

The AKE measurements of raters 1 and 2 from the first testing session were compared (Table 3). No significant difference between raters was noted for both lower limbs. The AKE test interrater reliability was excellent, with ICC$_{2,1}$ values of 0.87 (0.58–0.97; 95%CI) for the dominant knee and 0.81 (0.41–0.94; 95%CI) for the nondominant knee and standard error of measurement (SEM) values of 3.5° (18.0°–31.7°; 95%CI) and 3.8° (16.9°–31.7°; 95%CI) respectively. Minimal detectable change was between 9.7° and 10.5°. No significant difference in mean AKE measurements was noted between the first and second testing sessions for both raters (Table 4). The ICC$_{3,1}$ values ranged from 0.78 to 0.92.

DISCUSSION

Despite conflicting findings concerning the association of poor hamstring flexibility and the risk of hamstring muscle injury, assessment of hamstring flexibility is routinely conducted. Hamstring flexibility assessment is useful, especially in deciding an athlete's readiness to return-to-play following injury[11–13]. The AKE test is considered by some to be the gold standard for hamstring flexibility assessment[32].

With the aid of a simple stabilizing apparatus and a universal goniometer, we showed that a single assessor can conduct the AKE test easily. Interrater reliability ICC$_{2,1}$ values of 0.87 and 0.81 were found for the dominant and nondominant knees respectively. Further, the device also produced a standard error of measurement of 3.5° for the dominant knee and 3.8° for the nondominant knee. In addition, a good level of agreement between raters was established.

The test-retest reliability in this study was excellent,

with ICC$_{3,1}$ values ranging from 0.78 to 0.92 for both raters. Our finding is in agreement with earlier studies. Gajdosik reported a Pearson product-moment correlation coefficient for the AKE test of 0.99 for both lower extremities[33]. The higher reliability coefficient value could be explained by the different statistical analysis used. Further, the interval between the first AKE test and the retest session was very short. Both AKE tests were conducted on the same day 30 minutes apart, and this may have introduced systematic bias and affected reliability[2]. We chose an interval of one week between the test and retest AKE assessment because it reflects our clinical practice, as majority of cases are reviewed a week apart.

Using the mean value of AKE measurements from both knees, Gabbe reported excellent test-retest reliability, with ICC$_{3,1}$ values of 0.94–0.96[29]. In contrast to Gabbe et al., the current study evaluated each knee separately to explore any potential differences between the dominant and nondominant knees, and such differences in method may explain the wider ICC$_{3,1}$ values noted in this study. Further, the current study found no significant difference in AKE measurement between the dominant and nondominant knees of healthy individuals.

Similarly, a pilot study conducted by Davis reported excellent intrarater reliability of the knee extension angle (KEA) test, with an ICC$_{3,1}$ of 0.94[32]. The method of hamstring flexibility assessment employed by the previous author differs from that in the current study. In the former study, measurements were taken from two gravity inclinometers placed immediately superior to patella, and another was placed on the distal anterior tibia of the patient's lower extremity. In addition, the examiners performed hip flexion and knee extension passively, whereas participants

actively performed all movements in the current study.

Our finding concerning interrater reliability was consistent with those reported by earlier studies[29, 34]. Gabbe reported an interrater reliability $ICC_{2,1}$ of 0.93 and 4° SEM for the AKE test in 15 healthy participants of comparable age group (mean age: 31.6 years)[29]. Similarly, in a study to determine the effect of pelvic positioning and stretching method on hamstring flexibility, Sullivan reported an interrater reliability $ICC_{1,1}$ and SEM of 0.93 and 4.81° respectively for the AKE test among 12 healthy subjects[34]. On the other hand, Rakos et al. performed the AKE test with aid of an intricate stabilizing apparatus and demonstrated a good interrater reliability with an $ICC_{2,1}$ of 0.79 among children age 10 to 13 years old[22].

Despite demonstrating excellent interrater and test-retest reliability, a wide range of CIs was noted for some of the point estimates. Such findings, suggests that further research with larger samples may be necessary to determine the reliability estimates with greater precision[35]. Second, the reliability displayed in this study was based on assessment in healthy and uninjured volunteers. Whether the AKE test would be as reliable in a population of injured athletes remains unanswered. The AKE test might still be useful, especially during the preseason assessment, as most athletes are free from injury at the beginning of the season.

The current study demonstrated that the AKE test can be performed easily by a single assessor with a simple, portable, and inexpensive stabilizing apparatus and have excellent interrater and intrarater reliability.

REFERENCES

1) Cross KM, Gurka KK, Conaway M, et al.: Hamstring strain incidence between genders and sports in NCAA athletics. Athletic Train Sports Health Care, 2010, 2: 124–130. [CrossRef]

2) Askling C, Karlsson J, Thorstensson A: Hamstring injury occurrence in elite soccer players after preseason strength training with eccentric overload. Scand J Med Sci Sports, 2003, 13: 244–250. [Medline] [CrossRef]

3) Engebretsen AH, Myklebust G, Holme I, et al.: Intrinsic risk factors for hamstring injuries among male soccer players. A prospective cohort study. Am J Sports Med, 2010, 38: 1147–1153. [Medline] [CrossRef]

4) Gabbe BJ, Bennell KL, Finch CF, et al.: Predictors of hamstring injury at the elite level of Australian football. Scand J Med Sci Sports, 2006, 16: 7–13. [Medline] [CrossRef]

5) Verrall GM, Slavotinek JP, Barnes PG, et al.: Clinical risk factors for hamstring muscle strain injury: a prospective study with correlation of injury by magnetic resonance imaging. Br J Sports Med, 2001, 35: 435–439. [Medline] [CrossRef]

6) Burkett LN: Causative factors in hamstring strains. Med Sci Sports, 1970, 2: 39–42. [Medline]

7) Ekstrand J, Gillquist J: The avoidability of soccer injuries. Int J Sports Med, 1983, 4: 124–128. [Medline] [CrossRef]

8) Hennessey L, Watson AW: Flexibility and posture assessment in relation to hamstring injury. Br J Sports Med, 1993, 27: 243–246. [Medline] [CrossRef]

9) Jönhagen S, Nemeth G, Eriksson E: Hamstring injuries in sprinters the role of concentric and eccentric hamstring muscle strength and flexibility. Am J Sports Med, 1994, 22: 262–266. [Medline] [CrossRef]

10) van Beijsterveldt AM, van de Port IG, Vereijken AJ, et al.: Risk factors for hamstring injuries in male soccer players: a systematic review of prospective studies. Scand J Med Sci Sports, 2013, 23: 253–262. [Medline] [CrossRef]

11) Croisier JL, Forthomme B, Namurois MH, et al.: Hamstring muscle strain recurrence and strength performance disorders. Am J Sports Med, 2002, 30: 199–203. [Medline]

12) Drezner JA: Practical management: hamstring muscle injuries. Clin J Sport Med, 2003, 13: 48–52. [Medline] [CrossRef]

13) Mendiguchia J, Brughelli M: A return-to-sport algorithm for acute hamstring injuries. Phys Ther Sport, 2011, 12: 2–14. [Medline] [CrossRef]

14) Baltaci G, Un N, Tunay V, et al.: Comparison of three different sit and reach tests for measurement of hamstring flexibility in female university students. Br J Sports Med, 2003, 37: 59–61. [Medline] [CrossRef]

15) Booher JM, Thibodeau GA: Athletic Injury Assessment, 4th ed. McGraw-Hill, 1985.

16) Gajdosik R, Lusin G: Hamstring muscle tightness. Reliability of an active-knee-extension test. Phys Ther, 1983, 63: 1085–1090. [Medline]

17) Malanga GA, Nadler S: Musculoskeletal Physical Examination. Hanley & Belfus, 2006.

18) Bohannon RW: Cinematographic analysis of the passive straight-leg-raising test for hamstring muscle length. Phys Ther, 1982, 62: 1269–1274. [Medline]

19) Clark RA: Hamstring injuries: risk assessment and injury prevention. Ann Acad Med Singapore, 2008, 37: 341–346. [Medline]

20) Kuilart KE, Woollam M, Barling E, et al.: The active extension test and Slump test in subjects with perceived hamstring tightness. Int J Osteopath Med, 2005, 8: 89–97. [CrossRef]

21) Worrell TW, Perrin DH, Gansneder BM, et al.: Comparison of isokinetic strength and flexibility measures between hamstring injured and noninjured athletes. J Orthop Sports Phys Ther, 1991, 13: 118–125. [Medline]

22) Rakos DM, Shaw KA, Fedor RL, et al.: Interrater reliability of the active-knee-extension test for hamstring length in school-aged children. Pediatr Phys Ther, 2001, 13: 37–41. [Medline] [CrossRef]

23) Miyaguchi K, Demura S: Specific factors that influence deciding the takeoff leg during jumping movements. J Strength Cond Res, 2010, 24: 2516–2522. [Medline] [CrossRef]

24) Mutungi G, Ranatunga KW: Temperature-dependent changes in the viscoelasticity of intact resting mammalian (rat) fast- and slow-twitch muscle fibers. J Physiol, 1998, 508: 253–265. [Medline] [CrossRef]

25) Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1986, 1: 307–310. [Medline] [CrossRef]

26) de Weijer VC, Gorniak GC, Shamus E: The effect of static stretch and warm-up exercise on hamstring length over the course of 24 hours. J Orthop Sports Phys Ther, 2003, 33: 727–733. [Medline]

27) Portney LG, Watkins MP: Foundations of Clinical Research: Applications to Practice. 2nd ed. New Jersey: Prentice-Hall, 2000.

28) Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. Psychol Bull, 1979, 86: 420–428. [Medline] [CrossRef]

29) Gabbe BJ: Reliability of common lower extremity musculoskeletal screening tests. Phys Ther Sport, 2004, 5: 90–97.

30) Atkinson G, Nevill AM: Statistical methods for assessing measurement error (reliability) in variables relevant to Sports Medicine. Sports Med, 1998, 26: 217–238. [Medline] [CrossRef]

31) Macedo LG, Magee DJ: Differences in range of motion between dominant and nondominant sides of the upper and lower extremities. J Manipulative Physiol Ther, 2008, 31: 577–582. [Medline] [CrossRef]

32) Davis DS, Quinn RO, Whiteman CT, et al.: Concurrent validity of four clinical tests used to measure hamstring flexibility. J Strength Cond Res, 2008, 22: 583–588. [Medline] [CrossRef]

33) Gajdosik RL, Rieck MA, Sullivan DK, et al.: Comparison of four clinical tests for assessing hamstring muscle length. J Orthop Sports Phys Ther, 1993, 18: 614–618. [Medline]

34) Sullivan MK, Dejulia JJ, Worrell TW: Effect of pelvic position and stretching method on hamstring muscle flexibility. Med Sci Sports Exerc, 1992, 24: 1383–1389. [Medline]

35) Hopkins WG: Measures of reliability in sports medicine and science. Sports Med, 2000, 30: 1–15. [Medline] [CrossRef]