Original Research

Reliability of the Closed Kinetic Chain Upper Extremity Stability Test in young adults☆

Yslaíny Araújo Silva [a], Wbinayara Alves Novaes [a], Muana Hiandra Pereira Dos Passos [b], Vinicius Yan Santos Nascimento [b, *], Bruno Remígio Cavalcante [b], Ana Carolina Rodarti Pitangui [c], Rodrigo Cappato De Araújo [b, c]

[a] Student of the Graduate Course in Physiotherapy of the University of Pernambuco, Petrolina, PE, Brazil
[b] Associate Program of Post-graduation in Physical Education, UPE / UFPB, Recife, PE, Brazil
[c] Department of Physical Therapy, University of Pernambuco, Petrolina, PE, Brazil

A B S T R A C T

Objective: To analyze the reliability indicators of CKCUEST, and to investigate how many sessions are necessary to find a stable score.
Design: Reliability study.
Setting: Biomechanics laboratory.
Participants: The sample consisted of 36 healthy young adults, of both genders.
Main outcome measures: The subjects performed four CKCUEST with a minimum interval of three days between evaluations. The number of touches in each execution was counted, and then the average values of the number of rings, normalized score and power were calculated.
Results: The CKCUEST presented ICC ranging from 0.77 to 0.92 (mean number of touches), 0.80 to 0.94 (normalized score) and 0.91 to 0.98 (power). The Bland-Altman plots showed the presence of systematic error for all measures, as also confirmed by the ANOVA analysis. The scores increased every session for both genders. The boys' scores stabilized in the third session. However, the girls' scores did not stabilize.
Conclusion: The CKCUEST presents excellent reliability; however, there is the presence of systematic error between measurements. It is necessary to perform at least three sessions to obtain reliable data for male participants, and the female results were not conclusive.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Musculoskeletal evaluation of the upper extremity has been based on analysis of range of motion, strength and diagnostic tests (Tucci, Martins, Sposito, Camarini, & De Oliveira, 2014). In fact, these measurements provide information that can help in the individual evaluation, as well as in the follow-up of training or rehabilitation programs (Negrete et al., 2010). However, the inclusion of dynamic tests has been increasingly encouraged because they are complementary tools of low clinical cost that aggregate important information about the ability and functional performance (Roush, Kitamura, & Waits, 2007). In this context, the Closed Kinetic Chain Upper Extremity Stability Test (CKCUEST) can be an interesting option to evaluate the functional performance of the scapula and shoulder joint (De Oliveira et al., 2017; Tucci et al., 2014).

The CKCUEST is a simple, valid and low-cost test that can be easily applied in the clinical or sports context and which shows good correlation with measures of muscular strength (Lee & Kim, 2015; Roush et al., 2007). The test consists of an evaluation form of the function and stability of the upper limbs and has been used to identify the risk factors for shoulder pain (Tate et al., 2012), to determine the effectiveness of different types of intervention (Lust, Sandrey, Bulger, & Wilder, 2009) and to predict the evolution of throwing distances (Negrete, Hanney, Kolber, Davies, & Riemann, 2011).

Although several studies have found satisfactory reliability levels of CKCUEST scores in studies with a test-retest design

(Goldbeck & Davies, 2000; Tucci et al., 2014; Westrick, Miller, Carow, & Gerber, 2012), Oliveira et al. (De Oliveira et al., 2017) recently questioned the reliability levels of CKCUEST when only using one familiarization session. The authors identified moderate to excellent levels for the scores through the intraclass correlation coefficient (ICC). However, complementary analysis indicated the presence of systematic error, evidencing that there is a significant improvement in the retest, possibly explained by the familiarization effect to the task (De Oliveira et al., 2017).

In the literature it is similarly described that muscle strength assessment tests with maximal repetition calculations require more than a single familiarization session for stabilization and load definition (Hopkins, 2000). As CKCUEST has a strong correlation with the strength variable, it is possible to infer that the familiarization effect is present in its context, and that this effect may underestimate the individual's capacity and overestimate the responses resulting from training or rehabilitation programs.

Thus, the objective of this study was to investigate the reliability of the CKCUEST and to establish the number of sessions required to stabilize the scores.

2. Methods

2.1. Study design and participants

This is an reliability study conducted in the city of Petrolina (PE) with young adults of both genders. The sample calculation was performed following equations described by Shoukri, Asyali, Donner (2004), considering $\alpha = 0.05$; $\beta = 0.20$ (80% power); correlation ratio for null hypothesis (ρ H0) = 0.40, correlation ratio for alternative hypothesis (ρ H1) = 0.80, reaching a minimum required sample of 16 participants of each gender.

Participants were aged between 17 and 21 years, regular practitioners of some type of physical activity, being sports or resistance training, with minimum experience of two months, or those considered active according to the International Questionnaire of Physical Activity (IPAQ) (Matsudo et al., 2001). The exclusion criteria were related to non-attendance for the collections on the established dates or the impossibility of carrying out the CKCUEST. The final sample consisted of 36 participants, 18 male participants and 18 female participants, and their anthropometric characteristics are presented in Table 1.

The study was approved by the University's Ethics and Research Committee and all participants signed a free and informed consent form.

2.2. Procedures

Personal data (name, age and physical activity) and anthropometric measurements were initially performed. Then the volunteer was instructed about the positioning and execution of the Closed Kinetic Chain Upper Extremity Stability Test (CKCUEST) for male participants or CKCUEST modified for females participants.

The male participants adopted the push-up position, with their

feet positioned according to the height of the individual (75% of the height in meters) and demarcated in the ground by means of adhesive tape markers. The female participants performed a modified push-up with knee support on the ground, with the distance between the hands and the location for positioning their knees defined as 45% of height. The distance between the hands was 36 inches for all participants. During the test, the volunteer should remain with the trunk aligned, parallel to the ground, and their feet or knees positioned at the defined location (Tucci et al., 2014). The adopted positioning is demonstrated in Fig. 1.

The CKCUEST was carried out in four sessions on different days, with a minimum interval of three days and a maximum of seven days between evaluations. The test was performed three times in each session, with an interval of 45 s between repetitions. The subject remained in the push-up position while they alternatively touched their opposite hand for a period of 15 s. Each subject was allowed to perform the test once sub-maximally in order to familiarize themselves with the task prior to executing three repetitions of the actual test.

The presence of two evaluators was necessary, with the first evaluator being responsible for indicating the beginning and end of the test, as well as the verbal stimulus. Meanwhile, the second evaluator was positioned ahead of the participant, and was responsible for counting the number of touches. The evaluators were trained to perform the measurements prior to data collection.

Participants were not informed about the obtained scores in order to minimize the motivational effects. The average touches score was calculated on the basis of the mean number of touches recorded during the three attempts. The normalized score was calculated by dividing the average number of touches by the subject's height in meters. The power score was obtained by the product of the average number of touches and 68% of body weight in kilograms divided by 15 (Ellenbecker, Manske, & Davies, 2000).

2.3. Data analysis

The statistical procedures were conducted using two computer software packages: Statistical Package for Social Sciences (SPSS) version 20.0 and GraphPad Prism version 5.03. The distribution of data was analyzed by Shapiro-Wilk test. The comparison between the scores was accomplished using repeated measures ANOVA and Bonferroni post hoc.

The relative reliability analysis was calculated using the intraclass correlation coefficient (ICC3,3) two-way mixed model, with the absolute consent method and presentation of confidence interval 95% (CI 95%). ICC values above 0.75 were considered to represent excellent reliability; values between 0.40 and 0.74 represented moderate reliability; and values less than 0.40 indicated poor reliability (Fleiss, 1986). The Standard Error of Measurement (SEM) was calculated to estimate the variance of each score and the Minimal Detectable Change (MDC) was also evaluated to determine the threshold value for measurement error. Both were calculated by the respective equations:

$EPM_{95\%} = SD \cdot \sqrt{(1 - CCI_{test-retest})}$, for which the SD is the standard deviation of the mean of the first evaluation (test). $MDC = 1.96 \cdot EPM_{95\%} \cdot \sqrt{2}$, for which the constant 1.96 represents the z-score associated with the 95% confidence level.

The Bland-Altman plots were used to verify the absolute agreement between assessments from the scatter plot between the difference of the two assessments and the average of the two evaluations. Accounting for bias enabled determining the error limits of agreement, outliers, and trends.

Table 1
Mean and standard deviation of age and anthropometric characteristics of the total sample and stratified by gender (n = 36).

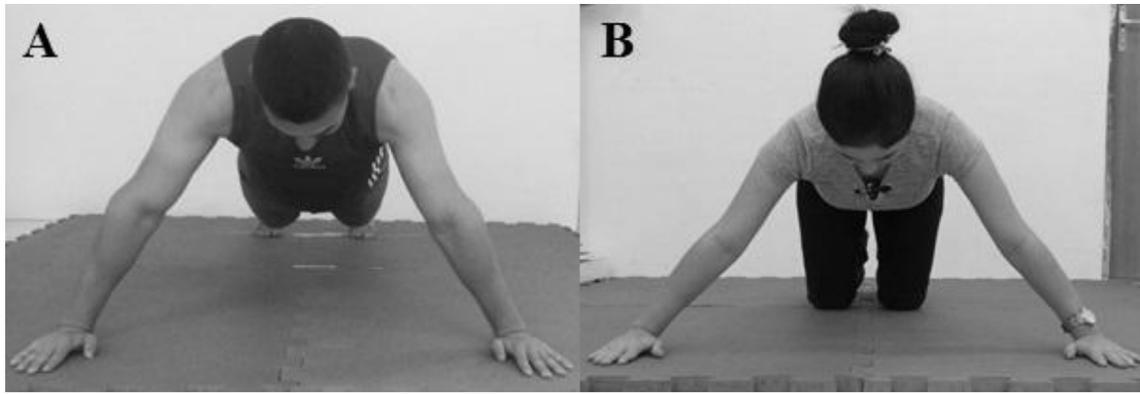| Variable	Total (n = 36)	Male (n = 18)	Female (n = 18)
Age (years)	19.5 (1.11)	19.50 (1.24)	19.50 (0.98)
Weight (kg)	65.72 (13.90)	74.73 (10.83)	56.70 (10.41)
Height (m)	1.66 (0.09)	1.72 (0.06)	1.60 (0.07)
BMI[a]	23.53 (3.19)	25.10 (3.06)	21.96 (2.52)

[a] BMI — Body mass index.

Fig. 1. Positioning of CKCUEST and CKCUEST modified. A - Position for male participants. B. Position for female participants.

3. Results

The mean and standard deviation data of number of touches, normalized score and power score of CKCUEST in the four evaluations are presented in Fig. 2. ANOVA showed a difference in the mean number of touches between the evaluations for males ($F = 19.355$, $p < 0.001$) and females ($F = 23.546$, $p < 0.001$). The normalized scores were different in the first three sessions, for males ($F = 18.246$; $p < 0.001$) and in the four sessions for females ($F = 22.227$; $p < 0.001$). In addition, there was a significant difference between the power scores for males ($F = 16.932$, $p < 0.0001$) and females ($F = 32.028$, $p < 0.0001$). This analysis indicates that the three variables have their values stabilized from the third session for male participants, while this stabilization only occurred in the power score in the female participants.

The ICC values demonstrate excellent reliability of the three variables in all assessments for both genders. The ICC values with their respective 95% confidence intervals, SEM, and MDC for each score of the CKCUEST are presented in Table 2.

Fig. 3 shows the Bland-Altman plots of the three variables analyzed in men. It is possible to observe the presence of systematic error between the first three evaluations, with negative bias and concentration of points below zero, indicating an increase in the scores at each session. However, stabilization of the scores seems to have occurred between the third and fourth sessions.

Fig. 4 shows the Bland-Altman plots of the three variables analyzed in women. It is possible to observe the presence of systematic error between all evaluations, with negative bias and concentration of points below zero, indicating an increase in the scores at each session.

4. Discussion

Although the test presents excellent values of relative reliability, absolute values indicate the presence of a systematic error, possibly related to the learning effect of the test, explaining the need for more than one evaluation to stabilize the scores. Three sessions are required for the male population, and in relation to the female scores it was not possible to precisely establish how many sessions are needed, as there was only stabilization in the values for the power variable in the fourth session.

Studies which have evaluated the reliability of the test have shown values varying from moderate to excellent reliability. One of the first studies developed on this subject was carried out with male university students, in which CKCUEST presented an ICC of 0.92 (Goldbeck & Davies, 2000). A more recent study conducted in the male and female adult population found an ICC of 0.97 (Lee & Kim, 2015). The test demonstrated excellent reliability values when reliability was assessed in symptomatic and asymptomatic subjects (ranging from 0.85 to 0.96 in the mean score of touches, 0.87 to 0.96 in the normalized scores and 0.82 to 0.96 in the power score) (Tucci et al., 2014). However, in the aforementioned studies it is possible to observe that the mean values found in the retest are higher than the values of the test. Although the authors do not present inferential statistics for this type of analysis, we speculate that perhaps there is systematic bias in this data that has not been investigated. In addition, the studies were limited to their interpretations as they were only based on ICC values.

Recently, Oliveira et al. (De Oliveira et al., 2017) assessed the reliability of CKCUEST in adolescent athletes, presenting an ICC of 0.68 for mean number of rings and normalized score, and ICC of
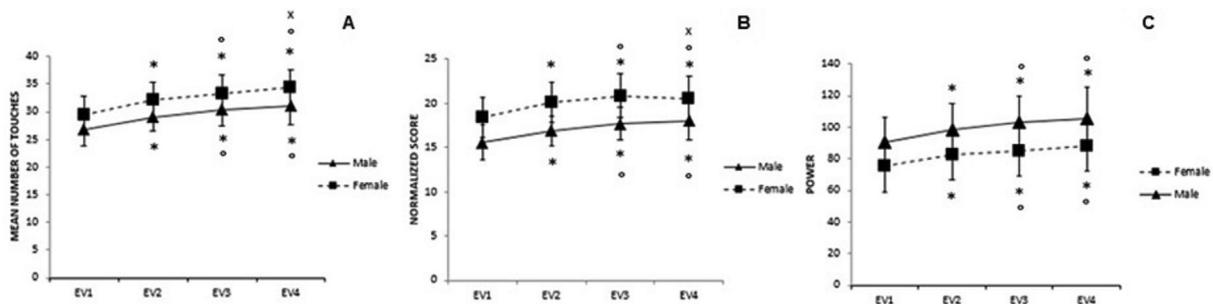

Fig. 2. Graphs presenting mean values and standard deviation of the CKCUEST variables stratified by sex. A- Mean number of touches; B– Normalized score; C– Power.
*$p < 0.05$; Ev1 vs Ev2, Ev1 vs Ev3, Ev1 vs Ev4.
□ $p < 0.05$; Ev2 vs Ev3, Ev2 vs Ev4.
□ $p < 0.05$; Ev3 vs Ev4.

Table 2
Intraclass Correlation Coefficient (ICC), Confidence Interval (95% CI), Standard Error of Measure (SEM) and Minimum Detectable Change (MDC) of mean number of touches, normalized score and power of the CKCUEST between evaluations.

|	Evaluation 1 x Evaluation 2			Evaluation 2 x Evaluation 3			Evaluation 3 x Evaluation 4		
	ICC (CI95%)	SEM	MDC	ICC (CI95%)	SEM	MDC	ICC (CI95%)	SEM	MDC
MALE									
Mean of touches	0.77 (−0.18−0.94)	1.22	3.39	0.87 (0.25−0.96)	1.05	2.92	0.80 (0.49−0.93)	1.10	3.04
Normalized score	0.82 (−0.13−0.95)	0.71	1.98	0.90 (0.33−0.97)	0.59	1.64	0.85 (0.61−0.94)	0.84	2.34
Power	0.91 (0.10−0.98)	4.92	13.63	0.96 (0.62−0.99)	3.34	9.27	0.93 (0.80−0.97)	5.17	14.33
FEMALE									
Mean of touches	0.77 (- 0.19−0.94)	1.53	4.25	0.91 (0.64−0.97)	0.99	2.75	0.92 (0.67−0.97)	0.93	2.58
Normalized score	0.80 (- 0.16−0.95)	1.02	2.83	0.93 (0.72−0.98)	0.65	1.79	0.94 (0.75−0.98)	0.61	1.69
Power	0.94 (0.16−0.99)	3.73	10.33	0.97 (0.89−0.99)	2.82	7.82	0.98 (0.91−0.99)	2.89	6.08
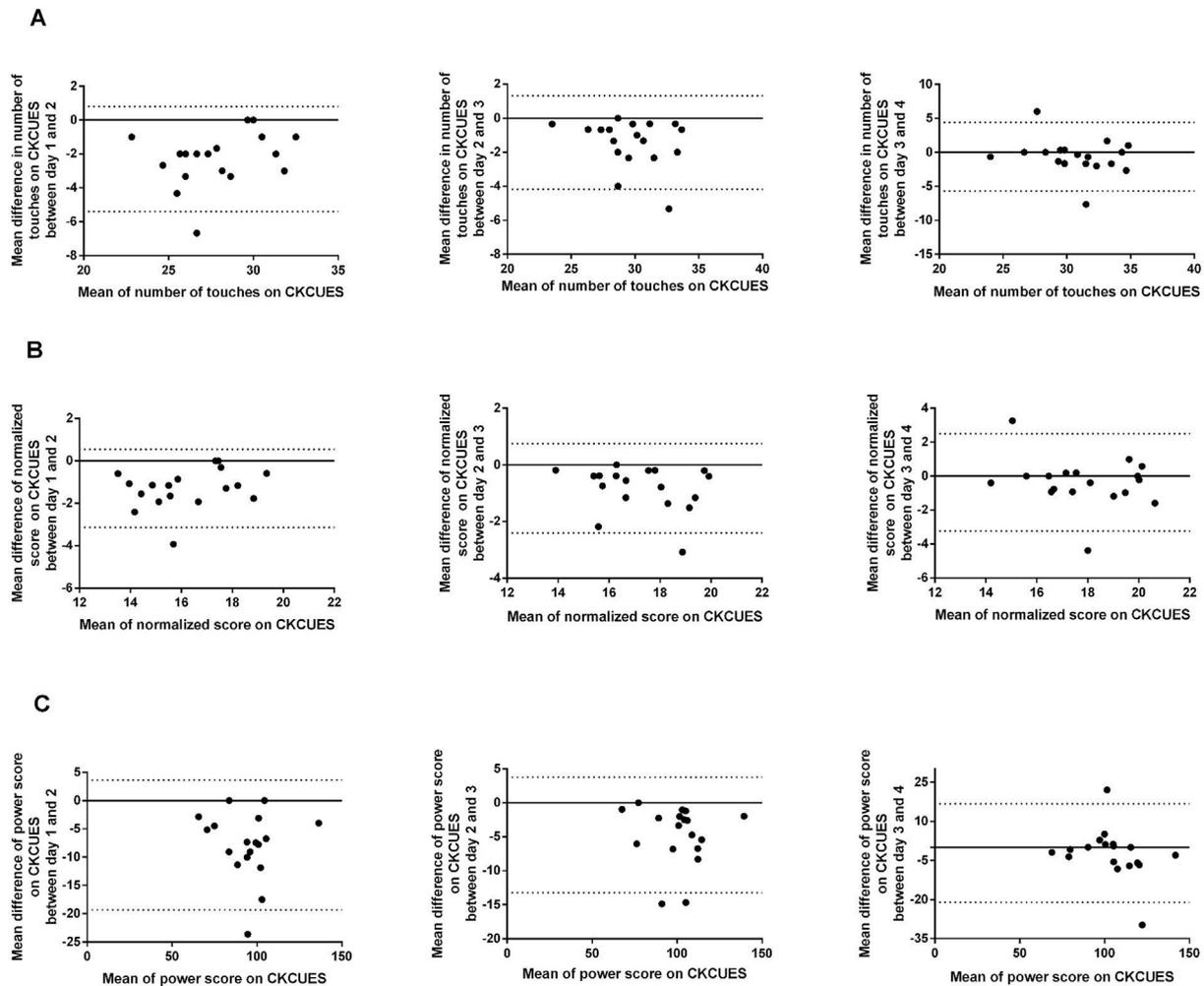
Fig. 3. Bland and Altman's graph of means dispersion of the variables of the CKCUES-test of the male participants. A - Mean number of touches; B— Normalized score; C - Power.

0.87 for power score. The authors further explored the Bland-Altman plot, in which they verified the presence of systematic error of measurement. Similarly, our results evidenced the presence of this systematic bias. In confronting the evaluations 1 and 2, and 2 and 3, it is possible to observe that the majority of the evaluated individuals presented negative values, indicating improvement in the scores. It is also worth mentioning the presence of outliers. Stabilization of these values was only observed in evaluations 3 and 4 for males for all scores. These data were also confirmed by ANOVA analysis.

Although the data presented excellent values for ICC, the mean scores increased significantly from one evaluation to another. These changes allow us to speculate that the process of learning and familiarization, which occurs more intensely in the first sessions, may influence the CKCUEST results, as was also observed by Oliveira et al. (De Oliveira et al., 2017) Following this reasoning, the use of the test in clinical practice may imply unreliable results. At the time of the test reassessment after application of a certain conduct, the evaluated individual may present higher score values that are not necessarily in response to the applied therapy, and may be due to the familiarization effect with the test.

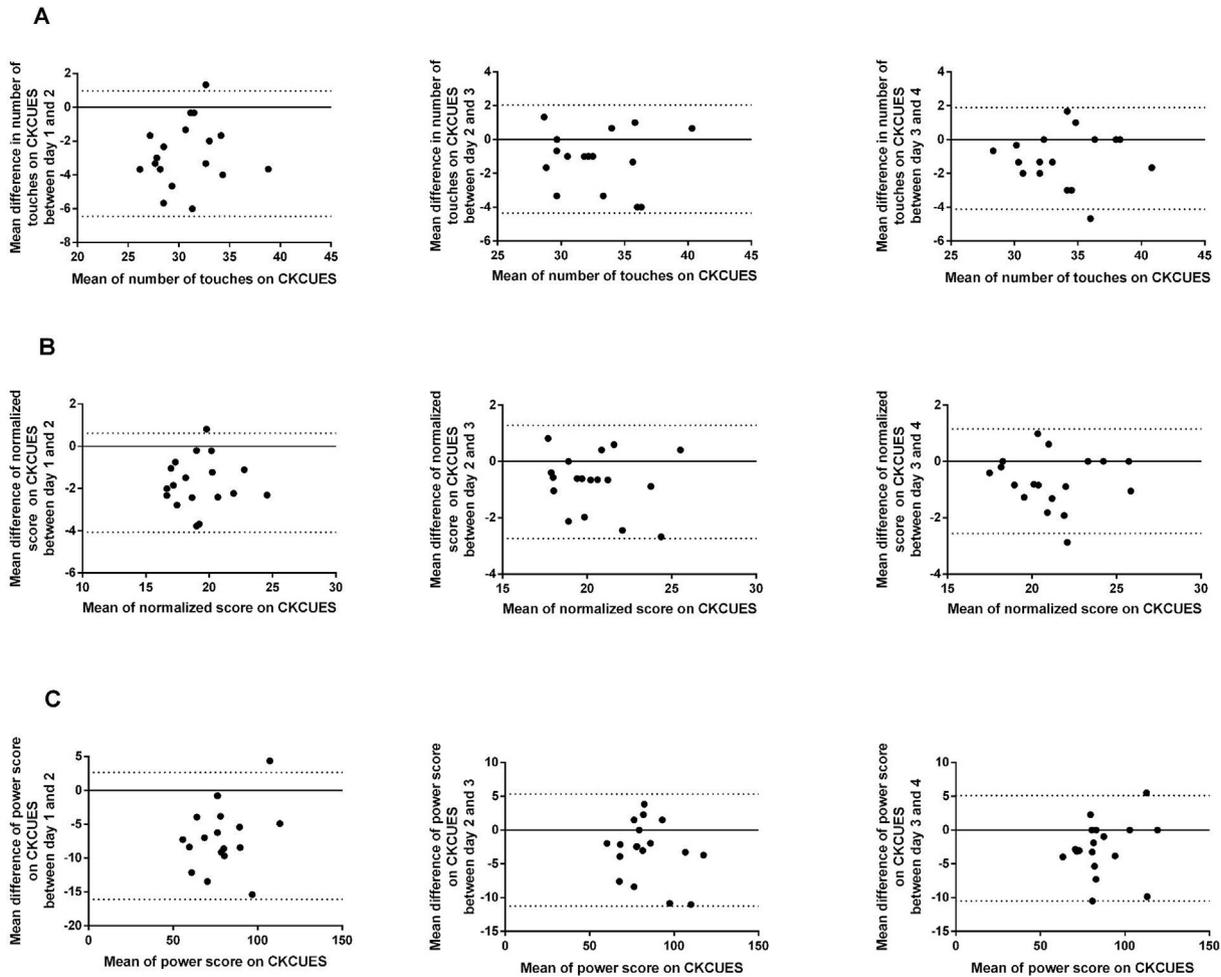The CKCUEST is a complex, multi-joint test which promotes

Fig. 4. Bland and Altman's graph of means dispersion of the variables of the CKCUEST of the female participants. A - Mean number of touches; B— Normalized score; C - Power.

weight bearing and requires motor control, stability and muscle strength. Lee and Kim (Lee & Kim, 2015) reported that the CKCUEST has a strong positive correlation with maximum grip strength and peak torque of internal and external rotation of the shoulder. Such facts may be related to the need to perform more than one evaluation for familiarization, as has already been suggested for maximal repetition tests in strength exercise (1-RM). (Dias, Ademar, Salvador, & Cyrino, 2009; Lima et al., 2013).

In male participants, three sessions were sufficient to stabilize the scores. However, the results for female participants indicate the need for more sessions to obtain reliable results. These findings can be partially explained by the type of previous experience of the subjects evaluated with the performed movement pattern or with the required muscle groups. In an empirical observation, Dias et al. (Dias et al., 2009) stated that women generally try to perform exercises involving lower limbs and glutes to the detriment of the trunk and upper limbs, different from the male population, which could explain the greater number of sessions for familiarization.

Another important aspect is the change of position to perform the test, which in female participants is performed with the support of their knees on the ground. This adaptation is characterized by a lower weight demand on the upper extremity when compared to the original position (Suprak, Dawes, & Stephenson, 2011). Tucci et al. (Tucci et al., 2017) describes that the average percentage of body weight sustained by the upper extremities in the modified push up, in a more precise calculation, would be of 24%. It is worth

mentioning that the percentage of body weight is part of the calculation of the power score, using the same value (68% of body weight) for men and women. This difference in values may also influence the results and analysis of this score.

Regarding the minimum detectable change data (MDC), it was shown that the minimum difference to be considered a real improvement for males is around three touches between the evaluations, corroborating the findings of Tucci et al. (Tucci et al., 2014) for this population. For women, the MDC value between the first and second sessions was greater than four touches. However, in the other comparisons the value of MDC approached the values of males. These findings make the learning effect for the female population more visible, which may be related to the fact that women train upper limb muscular groupings with a lower frequency or intensity when compared to the male population.

Another issue is that the CKCUEST studies do not standardize the distance to which the participants' feet or knees should be positioned, there is only research concerning the distance between the hands (Tucci et al., 2017). This lack of standardization makes the data inaccurate, since the axial load supported by a person during a push-up exercise changes according to the variations in the assumed position for the exercise (García-Massó et al., 2011; Koch, Riemann, & Davies, 2012; Tucci et al., 2017). In this way, load can be implemented or withdrawn during execution. To minimize these changes of positions, the positioning site in the present study was defined taking into account anthropometric characteristics of the

individual and demarcated in the ground so that all executions were performed with the same positioning.

The rigor in relation to the positioning to perform the test consists of a differential and strength of this study. However, the study participants were healthy individuals, thus the obtained results cannot be generalized to the symptomatic population. In addition, the four evaluations performed were not sufficient to determine the number of sessions required to stabilize the female scores, and therefore further studies are required.

5. Conclusion

The CKCUEST presents excellent relative reliability values for all scores, however there is systematic error possibly related to the learning effect. In this way, the results obtained in the first evaluations may underestimate the capacity of the individuals. The actual values for the male population are those found in the third session. The results were not conclusive for the female population, because they used the modified test and it did not show adequate reliability, so it may not be interesting for women to use this modified test.

Conflicts of interest

The authors declare no conflict of interest.

Ethical approval

Approval Opinion of the Research Ethics Committee of the University of Pernambuco - UPE, protocol: CAAE 38321114.0.0.0000.5207.

Ethical statement

The study was approved by Research Ethics Committee of the University of Pernambuco - UPE, protocol: CAAE 38321114.0.0.0000.5207 and all participants signed a free and informed consent form.

Funding

No funding was provided.

Acknowledgements

We thank the volunteers for their availability of time and willingness to participate in the collection procedures.

Appendix ASupplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ptsp.2019.04.004.

References

De Oliveira, V. M. A., Pitangui, A. C. R., Nascimento, V. Y. S., Da Silva, H. A., Dos Passos, M. H. P., & De Araújo, R. C. (2017). Test-retest reliability of the closed kinetic chain upper extremity stability test (CKCUEST) in adolescents: Reliability of CKCUEST in adolescents. *Int J Sports Phys Ther, 12*(1), 125–132.

Dias, R. M. R., Ademar, A., Salvador, E. P., & Cyrino, E. S. (2009). Familiarização ao teste de 1-RM em mulheres com experiência prévia em treinamento com pesos. *Rev da Educ Física/UEM., 20*(3), 423–429.

Ellenbecker, T. S., Manske, R., & Davies, G. J. (2000). Closed kinetic chain testing techniques of the upper extremities. *Orthop Phys Ther Clin North Am, 9*, 219–229.

Fleiss, J. L. (1986). *The design and analysis of clinical experiments.* Chichester, NY: Wile.

García-Massó, X., Colado, J. C., González, L. M., et al. (2011). Myoelectric activation and kinetics of different plyometric push-up exercises. *Jounal Strenght Cond Res, 25*(7), 2040–2047.

Goldbeck, T. G., & Davies, G. J. (2000). Test-retest reliability of the closed kinetic chain upper extremity stability test: A clinical field test. *Journal of Sport Rehabilitation, 9*, 35–45.

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

Koch, J., Riemann, B. L., & Davies, G. J. (2012). Ground reaction force patterns in plyometric push-ups. *The Journal of Strength & Conditioning Research, 26*(8), 2220–2227.

Lee, D. R., & Kim, L. J. (2015). Reliability and validity of the closed kinetic chain upper extremity stability test. *Journal of Physical Therapy Science, 27*(4), 1071–1073.

Lima, P. F. M., Araújo, R. C., Farah, B. Q., Cavalcante, B. R., Santos, M. A. M., & Ritti-Dias, R. M. (2013). Reprodutibilidade do teste de uma repetição máxima em exercícios de força com pesos livres Multiples sessions of one-repetition. *Rev Bras Atividade e Saúde., 18*(3), 378–380.

Lust, K. R., Sandrey, M. A., Bulger, S. M., & Wilder, N. (2009). The effects of 6-week training programs on throwing accuracy, proprioception, and core endurance in baseball. *Journal of Sport Rehabilitation, 18*(3), 407–426.

Matsudo, S., Araújo, T., Matsudo, V., et al. (2001). Questionário internacional de atividade física (IPAQ): Estudo de validade e reprodutibilidade no brasil. *Atividade física e Saúde., 6*(2).

Negrete, R. J., Hanney, W. J., Kolber, M. J., Davies, G. J., Ansley, M. K., McBride, A. B., et al. (2010). Reliability, minimal detectable change, and normative values for tests of upper extremity function and power. *The Journal of Strength & Conditioning Research, 24*(12), 3318–3325.

Negrete, R. J., Hanney, W. J., Kolber, M. J., Davies, G. J., & Riemann, B. (2011). Can upper extremity functional tests predict the softball throw for distance : A predictive validity investigation. *Int J Sports Phys Ther, 6*(2), 104–111.

Roush, J. R., Kitamura, J., & Waits, M. C. (2007). Reference values for the closed kinetic chain upper extremity stability test (CKCUEST) for collegiate baseball players. *North Am J Sport Phys Ther, 2*(3), 159–163.

Suprak, D. N., Dawes, J., & Stephenson, M. D. (2011). The effect of position on the percentage of body mass supported during traditional and modified push-up variants. *J Strenght Cond Res, 25*(2), 497–503.

Tate, A., Turner, G. N., Knab, S. E., Jorgensen, C., Strittmatter, A., & Michener, L. A. (2012). Risk factors associated with shoulder pain and disability across the lifespan of competitive swimmers. *Journal of Athletic Training, 47*(2), 149–158.

Tucci, H. T., Felicio, L. R., McQuade, K. J., Bevilaqua-Grossi, D., Camarini, P. M. F., & Oliveira, A. S. (2017). Biomechanical analysis of closed kinetic chain upper extremity stability test. *Jounal Sport Rehabil, 26*(1), 42–50.

Tucci, H. T., Martins, J., Sposito, G. D. C., Camarini, P. M., & De Oliveira, A. S. (2014). Closed kinetic chain upper extremity stability test (CKCUES test): A reliability study in persons with and without shoulder impingement syndrome. *BMC Musculoskeletal Disorders, 15*, 1.

Westrick, R. B., Miller, J. M., Carow, S. D., & Gerber, J. P. (2012). Exploration of the y-balance test for assessment of upper quarter closed kinetic chain performance. *Int J Sports Phys Ther, 7*(2), 139–147.