

# Linköping University Post Print

## **Intraobserver and interobserver reliability for the strength test in the Constant-Murley shoulder assessment**

Kajsa Johansson and Lars Adolfsson

N.B.: When citing this work, cite the original article.

Original Publication:

Kajsa Johansson and Lars Adolfsson, Intraobserver and interobserver reliability for the strength test in the Constant-Murley shoulder assessment, 2005, Journal of Shoulder and Elbow Surgery, (14), 3, 273-278.

<http://dx.doi.org/10.1016/j.jse.2004.08.001>

Copyright: Elsevier Science B.V., Amsterdam

<http://www.elsevier.com/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-14392>

**Intra- and inter-observer reliability for the strength test in the Constant-Murley shoulder assessment**

Kajsa M. Johansson<sup>1</sup>, RPT

Lars E. Adolfsson<sup>2</sup>, associate professor

<sup>1</sup> Department of Health and Society, Primary Care, Linköpings universitet, Sweden

<sup>2</sup> Department of Neuroscience and Locomotion, Orthopaedics and Sports Medicine, Linköpings universitet, Sweden

Corresponding author:

Kajsa M. Johansson

Department of Health and Society, Primary Care

Linköpings universitet

S-581 83 Linköping, Sweden

Phone: +46 13 22 10 33

Fax: +46 13 22 40 20

[Kajsa.Johansson@ihs.liu.se](mailto:Kajsa.Johansson@ihs.liu.se)

Source of support: Funding from the Faculty of Health Sciences, Linköpings universitet

## **Abstract**

This study evaluates the standardised strength test in the Constant-Murley-shoulder assessment on shoulder-healthy adults in a randomised, single-blind design. The following questions were to be answered; 1) can the spring-balance and a digital dynamometer both yield the same result 2) intra- and inter-observer reliability of the strength test 3) is the strength test sensitive to change in technique or affected by calculation with mean or maximum values?

Ten test-persons were included in a comparison of the Handyscale (digital dynamometer) and the mechanical spring-balance for concurrent validity resulting in ICC-values ranging from 0.96 to 0.99. For intra- and inter-observer reliability two observers tested 20 test-persons with the Handyscale, retest after two weeks. Regardless of techniques during testing, this resulted in almost perfect agreement (ICC range 0.89-0.98).

The digital dynamometer can replace the conventional spring-balance. The standardised strength test in the Constant-Murley-shoulder-assessment is reliable in young shoulder healthy persons, independent of technique or if calculated with mean or maximum values.

**Introduction**

Most shoulder disorders are not self-limiting which raises the need for efficacious clinical management<sup>7,10</sup>. Well-performed randomised controlled trials are demanded in order to evaluate treatments, and the outcome measures used in research as well as in clinical practice should be reliable and valid. Constant and Murley<sup>9</sup> have developed a frequently used shoulder score, the Constant-Murley shoulder assessment (C-M score), often used to measure the outcome after shoulder surgery<sup>13,15</sup>. One part of this score is a strength test, first described by Moseley<sup>20</sup>, measuring maximal isometric strength in 90° abduction. This part has been discussed and criticised for lacking standardisation<sup>1,2,8</sup>. Bankes et al<sup>1</sup> also raised the question about the technique used and recommended a “pull-force” with a fixed spring balance. They suggested a standardised test position that consists of the subject standing with the arm in 90° of lateral elevation in the scapular plane, the elbow extended and the forearm pronated. To our knowledge no one has tested and retested intra- and inter-observer reliability of the C-M score strength test in this standardised position. Further, the equipment used appears to differ and digital dynamometers as well as mechanical dynamometers occur in different centres. If a digital dynamometer is to be used instead of a mechanical one, its concurrent validity must first be determined by comparing it to gold standard, a mechanical spring-balance. In order to use this strength test as part of clinical evaluation, reliability needs to be further tested for any equipment used, to avoid measurement bias<sup>14</sup>.

The purpose of this study was to evaluate the standardised strength test in the C-M score on shoulder-healthy adults. The three following questions were to be answered; 1) can the spring-balance and a digital dynamometer both yield the same result 2) what are the intra- and inter-observer reliability of the strength test measured with the Handyscale, a digital dynamometer 3) is the strength test sensitive to change in technique, "pull-force" or "resisted-force", or affected by calculation with mean or maximum values?

## **Materials and Methods**

The test-persons ( $n = 30$ ) in the study were adult volunteers. All were students at the medical or technical faculty of Linköping University. They were included if they had no history of present or previous shoulder and/or upper extremity problems. None of the test-persons had performed this strength test before. The test-leaders were senior physiotherapy students with experience of the testing situation and the role as a test-leader from a pilot study performed on ten test-persons (20 shoulders), not included in the present study.

The study was divided in two parts, one to test concurrent validity and one to test intra- and inter-observer reliability. In both parts the equipment(s) were complemented with a loose handle from a pulley-apparatus in order to standardise the grip.

### *Concurrent validity*

The Handyscale® (Handyscale®, Bonso Electronics, Hong Kong, China) was compared with gold standard, a mechanical spring-balance (Tatsumi International Ltd., Ibarakai-city, Osaka, Japan), in order to test concurrent validity. The Handyscale is a digital dynamometer, measuring a maximum of 15 kilograms with two decimals and an interval of 20 grams. The mechanical spring-balance measures a maximum of 20 kilograms with an interval of 2 hectograms.

Ten test-persons (20 shoulders), four men and six women participated. The mean age was 23 years (range 19 to 25 years). The test-leaders tested five persons each. A blinded randomisation, by drawing of lots, decided which technique, which instrument and which arm to start with as well as which one of the test-leaders that would perform the test. The test situation and instructions were standardised as well as warm-up exercises including a circle movement in front of the body, drawing circles with the elbows with the hands placed on the shoulders and lifting the shoulder girdle. Ten repetitions of each exercise were performed. The test-persons also practised the technique by one sub-maximal attempt.

The test-position was the one recommended by Bankes et al<sup>1</sup> with the arm in 90° of lateral elevation in the scapular plane, the elbow extended and the forearm pronated. The test-person faced forward and the feet were positioned in line with the shoulders. The elbow was kept extended and all were told to avoid leaning their upper body during the test. The position was controlled visually by the test-leader during the test. The test-persons were told that maximal effort was necessary and to continue to pull respectively hold until instructed to stop. In the pull-technique the test-persons pulled until no higher value was reached, but no longer than 6-7 seconds. The force to resist during the "resisted-force" was applied slowly by the test leader until a visible break (joint movement) had occurred. The highest value obtained, in each repetition before joint movement, was noted. The equipment's, were unsecured for the technique "resisted-force" and for the "pull-force" secured in part of a pulley-apparatus which was fixed to the wall. Calibration of equipments was made with a weight of three kilograms before start of each test occasion. The set-ups are illustrated in figure 1 and 2.

A test-series contained three repetitions of for example "resisted-force" with the mechanical spring-balance. Then the same was performed with the other arm before changing back to the first arm to perform the "pull-force". Both techniques, two test-series per arm, were then repeated with the other equipment. The test-leader counted down from three and than cheered with the word hold (three times) or pull (three times), depending on technique.

Within one test-series of three repetitions, the test-persons were instructed to rest the arm at the side between each repetition and between series until they felt ready to perform again. The time of rest, between each effort, was 10-20 seconds and there was a natural rest when changing sides as well as when changing equipment. No discomfort was reported. The results were recorded by the respective test-leader and placed in an envelope with a code for that test-person. The data maintained concealed until analysis.



Figure 1 Illustrates the set-up for the “resisted-force” technique

#### *Intra- and inter-observer reliability*

Twenty test-persons (40 shoulders) participated in this part. There were ten men and ten women with a mean age of 26 years (range 21 to 34 years).

The test-leaders were the same as in the validity part. The test-retest was performed with exactly two weeks interval, in the same room with the same equipment (the Handyscale) and



Figure 2 Illustrates the set-up for the “pull-force” technique

at the same time at day. The order of test-leader, which arm and what technique to start with, was randomised blindly by drawing of lots. The standardised test-situation, instructions, periods of rest as well as preparations were the same as those earlier described in the validity part and repeated at retest.

A test-series contained three repetitions of for example "resisted-force". Then this was performed with the other arm before changing back to the first arm to perform the "pull-force". Both techniques, two test-series per arm, were then repeated with the second test-leader that waited outside the room.

The results were recorded by the respective test-leader and placed in an envelope with a code for that test person at the first test session as well as for retest. The data maintained concealed until analysis.

#### *Data analysis*

For concurrent validity and intra- and inter-observer reliability, agreement was analysed with a repeated measure analysis of variance (ANOVA)<sup>23</sup> to calculate Intraclass Correlation Coefficients (ICCs) with 95% confidence intervals using the SPSS version 9.0 (SPSS Inc. Chicago, USA)<sup>11</sup>. A "two-way random effect model" was chosen since the sources of variation in the concurrent validity part were the strength values and the equipment's and in the study of intra- and inter-observer reliability the strength values and the test-leaders.<sup>24</sup>

The ICCs were calculated with mean values of three repetitions as well as with the maximum values. In this study an ICC value of >0.81 was considered almost perfect, 0.61-0.80 as substantial, 0.41-0.60 as moderate, 0.21-0.40 as fair and 0.0-0.20 was considered as a slight agreement<sup>17</sup>.

## **Results**

The evaluation of concurrent validity resulted in ICCs ranging from 0.96 to 0.99. An almost

perfect agreement between the Handyscale and the mechanical spring balance, independently of the technique used or if calculated with mean or maximum values. (Table 1)

*Table 1* Intraclass correlation coefficients with confidence intervals for concurrent validity, the Handyscale versus the spring balance. The intraclass correlation coefficients are presented in relation to technique, "pull-force" or "resisted-force", and if calculated with the mean of three repetitions or with the maximum values

	Mean	Max
Pull-force	0.96	0.96
CI <sup>a</sup>	0.91-0.99	0.91-0.99
Resisted-force	0.99	0.99
CI <sup>a</sup>	0.98-1.0	0.96-0.99

<sup>a</sup> CI = 95% confidence intervals

The evaluation of intra-observer reliability resulted in ICCs ranging from 0.94 to 0.98 for test-leader A and ICCs of 0.90 to 0.96 for test-leader B. There was almost perfect agreement independent of technique or if calculated with mean or maximum values (table 2). The inter-observer reliability was also almost perfect with ICCs ranging from 0.89 to 0.96 at the first test occasion and ICCs of 0.89 to 0.97 at retest. An almost perfect agreement independently of technique used or if ICCs were calculated with mean or maximum values (table 2).

There were no significant differences between the strength values from the dominant or non-dominant arm. All maximum values and the mean strength values (mean of three repetitions) from the validity part are presented in table 3. The strength values from the reliability part are presented in table 4. Respective value was calculated as a mean of all maximum/mean values from each test-leader at test and retest.

In the validity part, the maximum strength values ranged from 4.4 kilograms (kg) to 8.0 kg for women and from 6.0 kg to 14.1 kg for men. In the reliability part the maximum strength

values ranged from 3.7 kg to 9.6 kg for women and from 5.6 kg to 14.9 kg for men. The mean of three repetitions rendered slightly lower values in both the validity and the reliability part, see table 3 and 4.

*Table 2 Intraclass correlation coefficients with confidence intervals for intra- and inter- observer reliability The intraclass correlation coefficients are presented in relation to technique, "pull-force" or resisted force", and if calculated with the mean of three repetitions or with the maximum values*

	Intra observer reliability				Inter observer reliability			
	Test leader A		Test leader B		Test 1		Test 2 "retest"	
	Mean	Max	Mean	Max	Mean	Max	Mean	Max
Pull-force	0.98	0.98	0.96	0.95	0.93	0.96	0.89	0.97
CI <sup>a</sup>	0.95-0.99	0.96-0.99	0.91-0.98	0.91-0.98	0.87-0.96	0.91-0.98	0.80-0.94	0.94-0.98
Resisted-force	0.94	0.95	0.90	0.91	0.89	0.91	0.94	0.95
CI <sup>a</sup>	0.88-0.98	0.78-0.98	0.78-0.95	0.80-0.96	0.78-0.94	0.82-0.96	0.79-0.98	0.82-0.98

<sup>a</sup>CI = 95% confidence intervals

Table 3 *Strength measurements from the concurrent validity part. Respective value calculated as a mean of all individual maximum values and as a mean of all means from both test-leaders*

	Female	Male
<b>Handyscale</b> (pull-force in kg <sup>a</sup> )		
Max	5.8	10.8
range	4.4-7.2	6.0-14.1
Mean	5.5	10.2
range	4.2-6.9	5.0-13.6
<b>Handyscale</b> (resisted-force in kg <sup>a</sup> )		
Max	5.9	10.8
range	4.8-7.6	7.3-12.9
Mean	5.5	10.3
range	4.6-6.8	7.0-12.6
<b>Springbalance</b> (pull-force in kg <sup>a</sup> )		
Max	5.8	10.3
range	5.0-8.0	7.0-12.2
Mean	5.5	9.7
range	4.6-7.3	6.3-11.5
<b>Springbalance</b> (resisted-force in kg <sup>a</sup> )		
Max	5.8	10.9
range	4.9-7.0	9.0-12.4
Mean	5.5	10.5
range	4.8-6.9	8.2-11.9

<sup>a</sup> kg = kilogram

Table 4 *Strength measurements from the reliability part. Respective value calculated as a mean of all individual maximum values and as a mean of all means, from both test-leaders at test and retest*

	Female	Male
<b>Handyscale</b> (pull-force in kg <sup>a</sup> )		
Max	6.5	10.6
range	4.2-9.6	5.6-14.9
Mean	6.0	10.0
range	3.9-9.2	5.1-14.9
<b>Handyscale</b> (resisted-force in kg <sup>a</sup> )		
Max	6.3	10.2
range	3.7-8.8	6.6-14.9
Mean	5.8	9.5
range	3.2-8.1	6.1-14.6

<sup>a</sup> kg = kilogram

## **Discussion**

This study found the standardised strength test in the C-M score, measured with the Handyscale, to be both intra- and inter-observer reliable. The Handyscale yielded almost the same results as the spring-balance, a high concurrent validity.

Since many sources might lead to measurement error, risking to be misinterpreted as a change in outcome in clinic as well as in research<sup>14</sup>, the factors controlled for in this study and others to consider is important to discuss.

To fulfil the purpose, this study was divided into two parts using different test-persons in each part to avoid influence from experience. Individuals without history of upper extremity problems were chosen. If patients with shoulder disorders had been used instead, there had been a risk of variation due to pain interfering with the evaluation of variation due to test-leaders and/or equipment's<sup>3</sup>. The almost perfect agreement when testing 'shoulder healthy people' is a prerequisite when it is used for patients with shoulder pathology, which is the original purpose of the C-M score. However, this reliability is based on young individuals.

A careful standardisation of the test situation and instructions were carried out through this study in concordance with what has been discussed as important during strength testing<sup>4,22</sup>.

The two weeks interval from test to retest using the same room and equipment at the same time at day was important to evaluate intra- and inter-observer reliability without diurnal variation. The test-persons' both arms were tested in order to increase the number of tests enlarging the risk of variation. The time of exactly two weeks between test-retest was judged as appropriate<sup>25</sup>. Since the test involves a maximum performance, a shorter time to retest could result in influences from muscle stiffness or soreness as well as bias due to test-leaders remembering data. A number of three repetitions were chosen to avoid errors due to fatigue<sup>1,21</sup>. The time of rest between each effort within the test-series, was 10-20 seconds. This is based on earlier research<sup>12,16,19</sup> and no participant commented that they needed more time.

The mean age was rather low, but earlier reported normal values for different age categories using “resisted-force” in 90° abduction, resulted in no significant differences between most age categories with exception for the comparison between those in 20-30 and 60-70 years of age.<sup>6</sup> In other words, most of the abduction strength values found in younger ages seemed representative also for those up to 60 years of age. Contradictory, Hughes et al<sup>12</sup> reported decreasing isometric abduction strength by age. This difference could be explained by the use of different equipment, the later using a modified Cybex II dynamometer.

In this study a handle from a pulley-apparatus to standardise the grip was used. This was similar to the original description<sup>20</sup>, but differs from the study by Bankes et al<sup>1</sup> where a adjustable strap placed around the forearm just proximal to the radiocarpal joint was used. Sapega et al<sup>22</sup> also recommended this, since involving the wrist could be a bias if the test-person has a problem with for example tennis elbow or by other reasons lacking distal stability or grip strength. In contrary the use of the handle makes the force-test more alike a functional situation, but can be less valid as a pure measure of abduction strength. However, our strength values did not differ significantly from those by Magnusson et al<sup>19</sup> who reported isometric “resisted-force” in 90° abduction, using a handheld dynamometer not involving the wrist. They used participants in the same age category as in the current study, suggesting that the influence from the forearm muscles is of minor importance.

The intra- and inter-observer reliability for the Handyscale is almost perfect (ICC >0.80), independent of technique used or if ICCs were calculated with mean or maximum values. No other study evaluating these aspects for the standardised position<sup>1</sup> has been found. Some studies have evaluated intra-observer reliability for isometric abduction, but in different degrees of motion or with other equipment's. Burnham et al<sup>5</sup> evaluated intra-observer reliability for the “pull-force” in 75° of shoulder abduction. They reported ICCs ranging from 0.51 to 0.78. In a study by Kuhlman et al<sup>16</sup>, isometric strength was measured using a

computerised dynamometer and in 90° abduction. Their ICC for intra-observer reliability was 0.72. These lower ICCs might be explained by the used equipments or positions.

Conboy et al<sup>8</sup> found the intra- and inter-observer error to be low, but imprecise in repeated measurements, concluding that the CM-score should not be used in patient follow-up. They evaluated inter-observer error of the strength test in the original position, which is not clearly described or standardised. A reliability estimate of 9.880 for patients with impingement was reported. The different way of analysing reliability makes it hard to compare with our results. Bankes et al<sup>1</sup> recommended a fixed spring-balance (“pull-force”) since the maximum value from the “resisted-force” technique differed significantly for shoulder patients. They reported that the standardised strength test was stable between repetitions using the fixed spring balance. The present study found the latter to be reliable both over time and between observers independent of if the equipment was fixed or unsecured. This diversity could be explained by the use of different test-persons, patients or shoulder healthy. This questions strength testing as a reliable method when assessing patients since the results would rather reflect the level of pain than the actual muscle strength. Such a value could still be meaningful, but needs awareness when interpreting the strength part in the C-M score.

Only a few of the test-persons reached a value that corresponded to maximum available points in the strength part in the C-M score (table 3 and 4), where the normal reference is a 25-year old male<sup>9</sup>. This means that even in the age between 20 and 35, it is hard for shoulder-healthy test-persons to reach a C-M score of 100 points. The test-protocol involved more than ten measurements, but the strength values did not decrease with increased number of repetitions, so this could not be explain by fatigue.

An earlier study of normal values reported higher strength values measuring isometric strength with a hand-held dynamometer in 90° abduction<sup>6</sup>. The standardised test position with the arm in 90° of lateral elevation in the scapular plane, used in the current study, might have

resulted in more specific supraspinatus activation. Less involvement of the deltoid muscle, which could explain the lower strength values reported in our study.

The question if there is a difference if ICCs are calculated using mean or maximum values seemed relevant since both have been recommended with different motivations in the literature. Mean values have been recommended since repeated scores may regress toward a more representative measurement<sup>4,19</sup>. The uses of maximum values are recommended since it reflects the individuals' best performance, relevant if strength is defined as the maximum voluntary force<sup>4</sup>. Calculating ICC with the mean value of three repetitions has two more risks of variation than in the case of calculation with the maximum value. Despite this the ICCs in this study are almost perfect, giving the analyser a free choice. In the C-M score the use of maximum values seems the most appropriate since is constructed to give more points for higher figures and endeavour to reach the maximum total score of 100 points.

The value of a measuring isometric strength in a single position can be questioned, since it is unlikely to be functionally representative. For patients, especially those with subacromial pain, the test position could be painful<sup>3</sup>.

In practical aspects it was easier to read the values from the Handyscale and we recommend the secured equipment ("pull-force") since it was easier to perform. The lightweight portable dynamometer, Isobex 2.0 (Cursor AG, Bern, Switzerland), might also be an alternative, but in comparison with the Handyscale there is a price difference in favour of the later<sup>18</sup>.

In conclusion the standardised strength test in the C-M score can be performed with either a digital- or a mechanical dynamometer. It is intra- and inter-observer reliable in young shoulder-healthy persons, independent of the technique used or if calculated with mean- or maximum values.

## **Acknowledgments**

We would like to thank Helene Bergström, RPT, and Åsa Johansson, RPT, for performing the strength tests. Thanks also to Susanne Ertzgaard, RPT, MSc, for valuable comments on the manuscript.

## **References**

1. Bankes MJK, Crossman JE, Emery RJH. A standard method of shoulder strength measurement for the Constant score with a spring balance. *J Shoulder Elbow Surg* 1998;7:116-121.
2. Bankes MJK, Emery RJH. Correspondence. *J Bone Joint Surg Br* 1997;79B: 696.
3. Ben-Yishay A, Zuckerman JD, Gallagher M, Cuomo F. Pain inhibition of shoulder strength in patients with impingement syndrome. *Orthopedics* 1994;17: 685-688.
4. Bohannon RW. Testing isometric limb muscle strength with dynamometers. *Phys Rehabil Med* 1990;2:75-86.
5. Burnham RS, Bell G, Olenik L, Reid DC. Shoulder abduction strength measurement in football players: reliability and validity of two field tests. *Clin J Sports Med* 1995;5:90-94.
6. Bäckman E, Johansson V, Häger B, Sjöblom P, Henriksson K-G. Isometric muscle strength and muscular endurance in normal persons aged between 17 and 70 years. *Scand J Rehabil Med* 1995;27:109-117.
7. Chard MD, Satelle LM, Hazleman BL. The long-term outcome of rotator cuff tendinitis - review study. *Br J Rheumatol* 1988;27:385-389.
8. Conboy VB, Morris RW, Kiss J, Carr AJ. An evaluation of the Constant-Murley shoulder assessment. *J Bone Joint Surg Br* 1996;78B: 229-232.
9. Constant CR, Murley AHG. A clinical method of functional assessment of the shoulder. *Clin Orthop* 1987;214:160-164.

10. Croft P, Pope D, Zonca M, O'Neill T, Silman A. Measurement of shoulder related disability: results of a validation study. *Annals of Rheumatic Diseases* 1994;53:525-528.
11. Eliasziw M, Young LS, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: Using goniometric measurements as an example. *Phys Ther* 1994;74:777-788.
12. Hughes RE, Johnson ME, O'Driscoll SW, An K-N. Age-related changes in normal isometric shoulder strength. *Am J Sports Med* 1999;27:651-657.
13. Iannotti JP, Bernot MP, Kuhlman JR, Kelley MJ, Williams GR. Postoperative assessment of shoulder function: a prospective study of full-thickness rotator cuff tears. *J Shoulder Elbow Surg* 1996;5: 449-457.
14. Krebs DE. Measurement theory. *Phys Ther* 1987;67:1834-1839.
15. Kronberg M, Wahlström P, Broström L-Å. Shoulder function after surgical repair of rotator cuff tears. *J Shoulder Elbow Surg* 1997;6:125-130.
16. Kuhlman JR, Iannotti JP, Kelly MJ, Riegler FX, Gevaert ML, Ergin TM. Isokinetic and isometric measurement of strength of external rotation and abduction of the shoulder. *J Bone Joint Surg Am* 1992;74A:1320-1333.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
18. Leggin BG, Neuman RM, Iannotti JP, Williams GR, Thompson EC. Intrarater and interrater reliability of three isometric dynamometers in assessing shoulder strength. *J Shoulder Elbow Surg* 1996;5:18-24.
19. Magnusson SP, Gleim GW. Subject variability of shoulder abduction strength testing. *Am J Sports Med* 1990;18:349-353.
20. Moseley HF. *Shoulder lesions*. p.28, 2 ed. Edinburgh: Churchill Livingstone, 1972.
21. Murray MP, Gore DR, Gardner GM, Mollinger LA. *Shoulder motion and muscle*

- strength of normal men and women in two age groups. *Clin Orthop* 1985;192:268-273.
22. Sapega AA, Kelley MJ. Strength testing of the shoulder. *J Shoulder Elbow Surg* 1994;3:327-345.
  23. Shrout PE, Fleiss JL. Intraclass correlation: Uses in assessing rater reliability. *Psychol Bull* 1979;86:420-428.
  24. SPSS. Base 9.0, applications guide. Chicago: SPSS Inc., United States of America, 1999.
  25. Streiner DL, Norman GR. Health measurement scales: A practical guide to their development and use. 2 ed. New York: Oxford University Press Inc.; 1998. p. 104-127.