Neurourology and Urodynamics 28:295–300 (2009)

Face Validity and Reliability of the First Digital Assessment Scheme of Pelvic Floor Muscle Function Conform the New Standardized Terminology of the International Continence Society

M.C.P. Slieker-ten Hove,[1,2*] A.L. Pool-Goudzwaard,[2,3] M.J.C. Eijkemans,[4] R.P.M. Steegers-Theunissen,[1,5,6,7] C.W. Burger,[1] and M.E. Vierhout[8]

[1]*Department of Obstetrics and Gynecology, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands*
[2]*Cobra Research Institute, Amersfoort, The Netherlands*
[3]*Department of Neuroscience, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands*
[4]*Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, The Netherlands*
[5]*Department of Epidemiology, Erasmus University Medical Centre, Rotterdam, The Netherlands*
[6]*Department of Pediatrics/Pediatric Cardiology Erasmus MC, University Medical Centre, Rotterdam, The Netherlands*
[7]*Department Clinical Genetics, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands*
[8]*Department of Obstetrics and Gynecology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands*

Aims: To test the face validity and reliability of a new digital pelvic floor muscle function (PFMF) assessment scheme that was designed on the basis of the recently standardized terminology of the International Continence Society. **Methods:** Study participants comprised 41 women, age 18–85 years. Data on age and parity were obtained. Face validity of the new assessment scheme was tested by three senior and one junior pelvic physiotherapists, using the Delphi technique. PFMF of each woman was assessed four times by three specially trained pelvic physiotherapists. Examiners were blinded to parity and other findings. To test reliability, Kappa (K) was used for the dichotomous variables and Weighted Kappa (K_w) for the items with more than two categories. **Results:** Mean age of the women was 41 years (SD 10.5); 14 were nulliparous (34.1%), 6 primiparous (14.6%), and 21 multiparous (51.2%). The new assessment scheme showed satisfactory face validity and intra-observer reliability but low inter-observer reliability. **Conclusions:** The new assessment scheme based on the terminology of the ICS showed satisfactory face validity and intra-observer reliability. It can therefore be considered suitable for use in clinical practice. More detailed redefinition of the described outcome measures is necessary to improve the inter-observer reliability. *Neurourol. Urodynam. 28:295–300, 2009.* © 2008 Wiley-Liss, Inc.

Key words: assessment; pelvic floor; pelvic floor musculature; pelvic floor muscle function; standardization

INTRODUCTION

Reliable assessment of pelvic floor muscle function (PFMF) is of great clinical importance, because incorrect PFMF is associated with incontinence, voiding disorders, defecation disorders, and dyspareunia.[1–4]

PFMF can be assessed by digital palpation,[5] observation, needle EMG, pressure measurement and/or ultrasound.[6,7] Although many researchers consider that digital assessment is unreliable, it is the most commonly used technique in clinical practice.[4] In the past decade, many different digital palpation methods have been developed. The Brink score[8] and the Laycock PERFECT assessment Scheme[9] are the most widely used. The Brink score employs a 4-point scale to assess the contraction pressure, vertical displacement and endurance of squeeze. The Laycock PERFECT assessment scheme uses a 6-point scale to score strength and endurance, the number of repetitions and fast contractions. Although in general the two scales have acceptable intra-observer and inter-observer reliability,[10–12] Bo[10] showed that they are not suitable for research purposes.

According to the standardized terminology, recently proposed by the Pelvic Floor Clinical Assessment Group (PFCAG) of the International Continence Societey,[13] an important omission is assessment of the involuntary, reflex contractions of the PFM that stabilize the pelvic floor and the urethra during an increase in intra-abdominal pressure (IAP). It has been demonstrated that effective reflex contractions play a major role in the prevention of leakage when there is an increase in IAP on the urethra.[4,14–19]

In 2004, Devreese[4] introduced a new scale that covered more PFM items and included the co-contraction of the Transverse Abdominal Muscles (TrAM). However, this scale was not based on the standardized terminology of the ICS. Therefore, we can conclude that none of the current assessment scales are based on the standardized terminology and they lack attention to involuntary contractions of the PFM. To conform with the new terminology o f the ICS and to test involuntary contractions, a new digital palpation scale is required. In our study, we formulated a new digital assessment scale that was partly based on definitions of outcome measures of existing scales but incorporated the items recently proposed in the new ICS terminology. Our assessment scale was tested on face validity and reliability.

Conflicts of interest: none.
*Correspondence to: M.C.P. Slieker-ten Hove, MA, COBRA Research SOMT, Softwareweg 5, 3821 BN Amersfoort, The Netherlands. E-mail: m.slieker@somt.nl
Received 28 April 2008; Accepted 1 August 2008
Published online 17 December 2008 in Wiley InterScience (www.interscience.wiley.com)
DOI 10.1002/nau.20659

MATERIALS AND METHODS

The new ICS terminology has not yet been translated into clear outcome measures. Therefore, a Delphi scheme was used in five rounds to reach consensus on the contents of the new protocol. This Delphi scheme synthesizes judgments when an uncertainty exists and utilizes the operator's common sense and the level of expertise.[20] In a literature study, our research team focused on digital PFMF assessment scales and excluded all types of equipment (such as ultra sound, manometry and ElectroMyoGram (EMG), MRI, non-English articles, surgery and pelvic organ prolapse) and the following MESH terms were used: PFMF, assessment and evaluation.

The literature search yielded 15 articles in which digital PFMF assessment had been used. These were then scored on the frequency of use in research, education and clinical practice. The research team comprised four members: two tutors in pelvic floor physiotherapy with over 12 years of experience, one senior physiotherapist with 10 years of experience in clinical work and the supervision of pelvic floor physiotherapy trainees and one junior physiotherapist without any teaching experience, but with 3 years of practical experience. Consensus was reached on the most relevant and frequently used PFMF assessment scales: *the Brink score*,[11] *the PERFECT Scheme*[9] *and the assessment method described by Devreese*.[4,14] To reach consensus on the most optimal definition of the outcome measures in the ICS terminology, the assessment schemes were discussed by the research team. Thus, the palpation protocol used in this study was based on the standardized terminology and existing definitions. Table I shows an overview of the palpation protocol used in this study. The last three columns list the currently available assessments scales which were used to refine our outcome measures as far as the standardized terminology allowed us to. They also provided insight into new and existing assessment items and different scoring methods.

TABLE I. PFMF Visual Inspection and Palpation Protocol According the Standardized Terminology of the ICS[13] Plus Assessment Scores From Brink, Laycock, and Devreese

Visual inspection	Score	Description				Laycock	Brink	Devreese
Inward movement	Yes	Any inward movement of the perineum						1 cm
	No	No inward movement of the perineum						
	Downward	Any downward movement of the perineum						
Co-contraction visible	Yes	Any co-activity of muscles other than TrAb						+ abd muscles
	No	No co-activity of other muscles was visible						
Relaxation	Yes	Relaxation visible directly after instruction						
	No	Absent, partial, hesitant or delayed relaxation						
Perineal movement during coughing	Yes	Any downward movement of the perineum						1 cm
	No	No downward movement						
	Inward	Any visible inward movement of the perineum						1 cm
Incontinence	Yes	Any incontinence						
	No	No incontinence						
Perineal movement during straining	Yes	Any downward movement of the perineum						
	No	No downward movement						
	Inward	Any inward movement of the perineum						
Palpation								
Pain	Yes	Any pain, location left-right-anterior-posterior						
	No	No pain						
Urethral lift	Yes	Urethral lift palpable					4 points	
	No	No urethral lift palpable						
Levator closure	Yes	Levator closure movement palpable						
	No	No levator closure movement palpable						
Symmetry left right	Yes	Complete symmetry between left and right						
	No	No symmetry L < R etc.						
Symmetry anterior posterior	Yes	Complete symmetry between anterior and posterior						
	No	No symmetry ant >post etc.						
Voluntary contraction	Strong	Strong closing and lifting cranio-anterior movement palpable				6 points	4 points	4–6 points
	Normal	Closing and lifting cranio-anterior movement palpable						
	Weak	Short contraction, no closure palpable						
	Absent	No contraction						
Endurance	≥10	9–7	6–4	3–1	0	10 points	4 points	
Fast twitch	≥15	14–11	10–6	5–1	0	10 points		2 points
Voluntary relaxation	Complete	Direct beyond rest level						
	Partly	Direct relaxation until rest level						
	Absent	No relaxation palpable						
During coughing								
Involuntary contraction	Yes	Contraction of the PFM						Deep PFM
	No	No contraction of the PFM						
Movement perineum	Yes	Any downward movement of the perineum						Superficial PFM
	No	No downward movement of the perineum						
During straining								
Involuntary relaxation	Yes	PFM relaxation palpable						
	No	No PFM relaxation palpable						
	Paradox	PFM contraction						

Subjects

Over a 2-month period, 41 women with and without pelvic floor disorders were recruited by an advertisement in the university faculty and city of Rotterdam and from a private physiotherapy practice outside the hospital. In the advertisement, women with and without symptoms of pelvic floor disorders, like vaginal bulging, urinary or faecal incontinence, were invited to respond. No form of payment was offered. Women were eligible to participate if they were aged between 18 and 85 years, had no degenerative neurological diseases and were able to comply with the instructions (no dementia, sufficient knowledge of the Dutch language). After enrolment, the participants filled in a short questionnaire about age and parity. All the participants received basic information on PFMF and an extensive explanation of the assessment scheme. The women received a small present on completion.

Assessment

The new scheme assessed voluntary and involuntary PFMF by means of observation and palpation (see Table I). To reach consensus on the exact contents of the assessment scheme prior to the reliability study, three pelvic physiotherapists were trained to use the new assessment scheme by a fully qualified pelvic floor physiotherapist tutor. All three physiotherapists had extensive experience with pelvic floor disorders, but needed to be trained in this specific protocol (two senior physiotherapists and one junior physiotherapist).

To investigate inter-observer reliability, all the components of the assessment scheme were done separately by each of the three pelvic physiotherapists. During the assessment, the women were placed in the lithotomy position on the treatment table, with the knees flexed at 90°, moderate abduction and 35° supported hip flexion. The examiners used non-allergic gloves lubricated with water-based anti-allergen. All the women were asked to empty their bladder before the assessment.

Before starting the present large-scale assessment, the new protocol was tested by the three physiotherapists in a pilot study on six women of the included women. This showed that the terminology was still ambiguous and led to different scoring. A 6th Delphi round was necessary to enable the final fine-tuning before starting the present assessments (face validity).

The 41 women were assigned at random to the three examiners. Before the assessment, a pelvic physiotherapy student informed the participants about the assessment scheme and answered their questions. Each woman was tested in three different rooms and the examiners changed room every 10–15 min. In this way, each of the women was examined four times: the first three times in a random order by the three different physiotherapists and the fourth time by the physiotherapist who performed the first assessment.

All the examiners were blinded to each others' findings and to parity of the subjects. To enable the subjects to rest in between, the whole scheme of four assessments was spread over 60 min. After the last assessment, the examiner informed the woman about her PFMF.

The exact instructions are shown in Table II. During the assessment, all verbal instructions were posted on the wall behind the table to ensure that the physiotherapists gave exactly the same instructions in every test. Digital assessment was performed with one finger. Palpation with two fingers was used to measure the closing movement of the levator ani. Relaxation was tested after a contraction.

Inter-observer reliability of the assessments was expressed as the percentage of agreement between two observers, calculated as the number of agreements divided by the total number of assessments. Likewise, intra-observer reliability was expressed as the percentage of agreement between the two assessments made by the same examiner. To calculate whether agreement was beyond chance, the Kappa (K) statistic was used for dichotomous variables and the Weighted Kappa (K_w) for the items with more than two categories. The criteria given by Landis and Koch[21] (Table III) were used to express the strength of the agreement. The kappa values for inter-observer agreement between three examiners were estimated by means of the Intra-Class Correlation Coefficient (ICC), which is equivalent to a quadratic ally weighted kappa.[22]

When variables have very skewed distributions, e.e. if the vast majority of women are rated in one particularly category by two examiners, the level of agreement will be high, but as a large part of that agreement can be attributed to chance alone, the kappa value actually may be very low.[23] All the statistical analyses were performed with SPSS package 11.5.

The study was approved by the Medical Ethics Research Committee (METC) of the Erasmus Medical Centre in Rotterdam, the Netherlands. All subjects gave informed consent before they entered the study.

TABLE II. Detailed Verbal Instructions During the Assessment

Assessment item	Verbal instruction
Visual inspection	
Voluntary contraction	Lift and squeeze your pelvic floor Try to avoid loss of urine or flatus
Palpation	
Voluntary contraction	Lift and squeeze your PFM as hard as possible
Endurance	Make a steady but firm contraction and hold it as long as you can, while repeating "hold and hold and hold!"
Fast contraction	Make fast, short and strong contractions, while repeating contráct, contráct, contráct!
Reflex movement	
Cough	Cough forcefully
Push	Give a strong push

RESULTS

A total of 41 women were examined (mean age of 41 years (SD 10.5), range from 22 to 63 years). Age did not have a normal distribution, as assessed by the Kolmogorov–Smirnov test: *P* 0.025. Data on parity status showed that 14 women were nulliparous (34.1%), 6 primiparous (14.6%), and 21 multiparous (51.2%). Table IV presents the overall agreement and K_w of the inter-observer variability of all three examiners, together with the agreement and Kappa of the intra-observer variability.

TABLE III. Criteria for the Interpretation of Cohen's Kappa Following Landis and Koch[21]

Value of Kappa	Interpretation
<0	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

TABLE IV. Overall Findings of Agreement of PFM (Dys)Function (Agr.%/K) and K_w

	Intra-observer reliability			Inter-observer reliability		
	K, agr %	K_w	95% CI	K, agr %	K_w	95% CI
Visual inspection						
Inward movement	100	[a]	0	100	[a]	[a]
Co-contraction visible	**75**	**0.48**	0.20–0.69	**61**	**0.52**	0.34–0.69
Relaxation	97.6	[a]	0	**97.6**	**0.75**	0.62–0.85
Perineal movement during coughing	**70.7**	**0.54**	0.28–0.73	44	0.33	0.14–0.53
Incontinence	97.6	[a]	0	**97.6**	**0.75**	0.62–0.85
Perineal movement during straining	90.2	0.33	0.03–0.58	82.9	0.013	0.14–0.22
Palpation						
Pain	**89.7**	**0.79**	0.60–0.87	**89.5**	**0.85**	0.76–0.91
Urethral lift	95.1	[a]	0	82.9	0.08	−0.09–0.30
Levator closure	92.7	.39	0.10–0.62	**82.9**	**0.45**	0.26–0.63
Symmetry left right	**87.8**	**0.64**	0.42–0.79	62.5	0.16	−0.03–0.37
Symmetry anterior posterior	**89.2**	**0.68**	0.46–0.82	46.2	0.10	−0.08–0.32
Voluntary contraction	**75.6**	**0.67**	0.46–0.81	**56.1**	**0.64**	0.48–0.77
Endurance	**78**	**0.76**	0.60–0.86	36.6	0.37	0.17–0.56
Fast contraction	**61**	**0.60**	0.37–0.77	**39**	**0.47**	0.29–0.65
Voluntary relaxation	**87.8**	**0.76**	0.59–0.87	39	0.17	−0.01–0.38
Palpation during coughing						
Involuntary contraction	**82.9**	**0.66**	0.44–0.80	46.3	0.33	0.14–0.53
Movement perineum	**95.1**	**0.77**	0.61–0.87	75.6	0.03	−0.13–0.24
Palpation during straining						
Involuntary relaxation	**80**	**0.60**	0.36–0.79	61	0.15	−0.03–0.36

[a]The K_w measure could not be calculated because to few differences were observed and scoring results were dichotomized.
Values in bold ($P < 0.05$).

The intra-observer reliability was, according to the table of Landis and Koch (Table III), substantial for eight items (pain, two items of symmetry, voluntary contraction, endurance, voluntary relaxation, involuntary contraction and perineal movement during coughing). The inter-observer reliability of pain assessments was almost perfect. Substantial inter-observer reliability was found for visual inspection of relaxation after contraction and the presence of incontinence during coughing and for digital assessment of voluntary contractions. Overall, there was very little variability in the PFMF scores between the women.

DISCUSSION

This is the first study in which a PFMF assessment scheme has been designed on the basis of the new standardized terminology of the ICS. Definitions of dysfunction were transformed into refined outcome measures and the scheme was tested for reliability. Existing assessment schemes are all based on self-formulated definitions.

The PFCAG of the ICS, who is responsible for the new terminology on PFMF, stated that the new definitions are descriptive and meant for daily clinical practice. However, this terminology is vague and not yet well defined. Therefore, to implement this new terminology in an assessment scheme, we performed 6 Delphi rounds to refine the outcome measures and to establish face validity. Several items from existing protocols have been included in the new scheme (Table I). Some items are comparable with those in other schemes and others now have a different outcome measure.

Face Validity

Face validity was tested by fully trained pelvic floor physiotherapists alone (mono-disciplinary). Although their level of experience varied, higher face validity could have been reached for example by including gynaecologists in the team of (multidisciplinary) experts. Furthermore, only a few studies were suitable to support the expert opinions in the preparation of proper definitions. Development of the assessment was therefore troublesome, time-consuming and placed high demands on the participants in the Delphi rounds. The final result is the first complete assessment protocol in accordance with the ICS.

Intra-Observer Reliability

Intra-observer reliability was moderate to substantial, which is comparable with the other studies by Bo and Finckenhagen,[24] Brink, Devreese, and Laycock.

All the *visual inspection* items showed close agreement. This may have been due to the dichotomization of items and the low level of variability in PFMF scores between the women. For example, most of the women had a positive score on the item "seeing an inward movement." Observation of an inward movement has been studied earlier.[10,24] Bo demonstrated an inward movement of 10.8 mm (SD 6.0) during MRI in the sitting position and confirmed this later by ultrasound in the supine position.[25] Apparently, even in the new scheme, the definition does not yet make sufficient distinction. Two items had only moderate intra-observer reliability: visual inspection of co-contractions and perineal movement during coughing (0.48 and 0.54). Probably, fatigue played a role in the 4th examination. Although we performed the four assessments over the period of 1 hr to enable the women to rest as much as possible, we cannot be sure that fatigue did not occur. Until now, it is unclear how 4 consecutive tests within 1 hr affect PFMF. In the scheme developed by Devreese et al.[4] "perineal movement during coughing" scored the abdominal wall activity and perineal activity, which resulted in 1 cm of movement. However, this item does not form part of the ICS terminology. Therefore we excluded abdominal

wall activity from our test on perineal movement during coughing.

Relaxation (after a voluntary contraction), inward movement and perineal movement during straining showed high agreement, weighted Kappas could only be calculated on the straining part of the scheme. It was found that relaxation had high intra-observer reliability agreement and formed a good measure. As the variability within the group was low, it was impossible to calculate weighted Kappas of the intra-observer reliability scores.

PFM strength during voluntary contraction was tested using *palpation* (in line with the regular method), and classified as absent, weak, normal and strong, just like in the Brink score. Palpation showed substantial reliability with a K_w of .67 for intra-observer reliability. This is in agreement with the study by Bo and Finckenhagen,[26] although they stated that categorizing the results on 3- or 5-point scales may not have sufficient sensitivity to differentiate between individuals. This was also confirmed in the studies by Fitzgerald et al.[19] and Frawley et al.[27] However, we followed new definitions proposed by the ICS and would like to emphasize that strength measurement is only one aspect of PFMF and as such, only forms a small part of the whole assessment scheme.

Endurance showed substantial (0.76) intra-observer reliability. This was in conformity with the study by Bo and Finckenhagen,[22] but disagreed slightly with the study by Devreese et al.[4] It is possible that this difference was caused by the study population of Devreese, which comprised continent and incontinent women, that is, optimal and minimal PFMF. In our group PFMF did not show any great variability. To test endurance we scored the quality of a voluntary contraction during 10 sec in the same way as Devreese et al.[4] and Laycock in the PERFECT scheme. In contrast, the Brink score only uses a maximum of 3 sec. It is our opinion that 10 sec are necessary to test endurance properly.

Fast contractions were also tested in conformity with existing methods and showed moderate agreement (0.60). This outcome is difficult to compare to other studies, because different methods were used. In the study by Devreese et al.,[4] only the contraction speed was measured, and not the number of fast contractions. We proposed to assessing fast contractions in the new scheme, based on our hypothesis that in daily life, frequent fast contractions are necessary to coordinate well-timed reflex contractions. Thus, the number of contractions is also important.

Pain showed substantial intra-observer reliability (0.79). This item was not scored in earlier PFM assessments. Pain can influence PFMF[28] and the high K_w in our study proved the high reliability. We regard information about pain to be essential in the assessment of PFMF.

Palpation of asymmetry showed substantial (0.64–0.68) intra-observer reliability. This test is important, because Dietz[29] has recently found that it can help to diagnose trauma in the levator ani. They suggested to add strength asymmetry measurements to PFMF assessment schemes. In our assessment scheme, asymmetry was tested in a voluntary contraction, without asking for maximal strength. Based on the findings of Dietz et al.[29] we recommend the addition of asymmetry assessments, especially strength measurements, to gain insight into the possible presence of levator trauma.

Closing of the levator was also included, but intra-observer reliability was low (0.39), whereas overall agreement was 92.7%. This was probably due to fatigue during the later measurements. Fatigue is known to influence total PFMF and as stated above, strength is not the only function that is responsible for good and adequate PFMF.[4,30,31]

Urethral lift during a contraction showed high agreement, but no K_w. Our decision to score any sign of upward lift as positive was probably the cause of this result. Urethral lift is tested in 4 items in the Brink-score and demonstrated a moderate reliability (0.51).[11] This point could be taken into account when the PFCAG make adjustments to PFMF terminology in the future.

Relaxation after a contraction showed high intra-observer reliability (0.76). However, the scoring of relaxation according the ICS is disputable, because they only define three steps: complete (relaxation beyond the resting level), partial (relaxation until the resting level) and absent (no relaxation palpable). We feel that incomplete relaxation is missing form the terminology (i.e., relaxation that does not (quite) reach the resting level). This point caused a great deal of discussion between our examiners. In clinical practice, we are used to scoring relaxation on the three ICS items and on our own "incomplete relaxation" item, which might be the reason for the low intra-observer reliability scores.

During coughing, we tested involuntary, reflex contractions of the pelvic floor muscles by looking for inward movements and urethral lift. Furthermore, we tested involuntary relaxation during straining by looking for downward movements of the perineum. It is important to test the item of involuntary contraction defined in the ICS terminology, because downward movement of the perineum and incontinence during a sudden increase in IAP (not to be confused with straining for defecation) are the result of inadequate contraction of the pelvic floor muscles.[14,15,32,33] The scoring of involuntary movements during coughing and straining showed substantial intra-observer reliability. Our reflex contraction results seem to be in line with Devreese et al.[4] whose the intra-observer reliability was almost perfect (0.88). However, we wish to in to emphasize that in general good voluntary contraction seen in the assessment that results in positive urethral lift, is no guarantee that effective reflex contractions always occur when needed.[4,14] The outcomes of these items can form helpful indicators in the development of a functional training program.

Inter-Observer Reliability

Inter-observer reliability was generally disappointing. Although we endeavored to avoid any bias in the assessments, we were unable to achieve reliability on many of the items. Fatigue may form an important explanation for the changes in PFMF during the test scheme and the low agreement scores. Other explanations might be a learning effect by the women and inter-tester differences. Although the verbal instructions were standardized, the tone and personality of the examiner can interfere also.

Substantial agreement (0.75) was seen in the visual inspection part of the scheme that tested relaxation after a contraction and the symptom of urinary incontinence. Apparently, the contrasts in function are clearly visible. A few items on the scheme were good: pain (0.85) and a voluntary contraction (0.64), while two items were moderate: fast contractions (0.47) and the levator closure (0.45).

As pain has never been scored in the earlier studies on PFM assessments, our results cannot be compared to the literature. This also applies to levator closure that has not been tested before. We believe that this item also needs further refinement to increase inter-tester reliability. Due to different testing methods and different outcome measures, not all of our reliability scores of voluntary contractions and fast contractions are comparable with other studies.

Discrepancies Between Intra-Observer Reliability and Inter-Observer Reliability

There were large differences in intra-observer and inter-observer reliability on the outcome symmetry, with high scores on intra-observer reliability (K_w 0.64 and 0.68) and low scores on inter-observer reliability (K_w 0.16 and 0.10). An explanation probably lies in different interpretations by the three examiners. Therefore, this item also needs more refinement and/or a different approach.

Palpation of the involuntary contraction of the pelvic floor muscles and movement of the perineum showed substantial intra-observer reliability (0.66), but only moderate inter-observer reliability (0.33). Therefore, these items need further refinement.

This study is an important step forward in the definition of outcome measures to assess PFMF, but it has its limitations. Our study group appeared to be less heterogeneous than would have liked. Age did not have a normal distribution, but as none of our statistical analyses required the assumption of normality of age, this characteristic was irrelevant. The definitions of the ICS need further refinement before they can be incorporated into an assessment scheme that is also suitable for research purposes.

CONCLUSION

Our new PFMF assessment scheme was designed according to the standardized terminology of the ICS. Based on the moderate to substantial intra-observer reliability and face validity, it can be considered reliable in clinical practice. Before implementation for research purposes, the outcome measures need detailed refinement in view of the disappointing inter-observer reliability.

ACKNOWLEDGEMENTS

P.M.M. van Nierop, MSPT, J. Dijkstra-Eshuis, MSPT and J. Rodenburg, PFPT for their contribution to the study.

REFERENCES

1. Eliasson B, Shukla PK. Dynamics of whistler spherom aks in magnetized plasmas. Phys Rev Lett 2007;99:205005.
2. Lagro-Janssen TL, Debruyne FM, Smits AJ, et al. Controlled trial of pelvic floor exercises in the treatment of urinary stress incontinence in general practice. Br J Gen Pract 1991;41:445–9.
3. Bo K, Talseth T, Holme I. Single blind, randomised controlled trial of pelvic floor exercises, electrical stimulation, vaginal cones, and no treatment in management of genuine stress incontinence in women. Br Med J 1999;318:487–93.
4. Devreese A, Staes F, De Weerdt W, et al. Clinical evaluation of pelvic floor muscle function in continent and incontinent women. Neurourol Urodyn 2004;23: 190–7.
5. Bo K, Sherburn M. Evaluation of female pelvic-floor muscle function and strength. Phys Ther 2005;85:269–82.
6. Dietz HP. Why pelvic floor surgeons should utilize ultrasound imaging. Ultrasound Obstet Gynecol 2006;28:629–34.
7. Thompson JA, O'sullivan PB. Levator plate movement during voluntary pelvic floor muscle contraction in subjects with incontinence and prolapse: A cross-sectional study and review. Int Urogynecol J Pelvic Floor Dysfunct 2003;14: 84–8.
8. Brink CA, Sampselle CM, Wells TJ, et al. A digital test for pelvic muscle strength in older women with urinary incontinence. Nurs Res 1989;38:196–9.
9. Laycock J. Incontinence. Pelvic floor re-education. Nursing (Lond) 1991;4:15–7.
10. Bo K. Pelvic floor muscle strength and response to pelvic floor muscle training for stress urinary incontinence. Neurourol Urodyn 2003;22:654–8.
11. Brink CA, Wells TJ, Sampselle CM, et al. A digital test for pelvic muscle strength in women with urinary incontinence. Nurs Res 1994;43:352–6.
12. Hundley AF, Wu JM, Visco AG. A comparison of perineometer to brink score for assessment of pelvic floor muscle strength. Am J Obstet Gynecol 2005;192: 1583–91.
13. Messelink B, Benson T, Berghmans B, et al. Standardization of terminology of pelvic floor muscle function and dysfunction: Report from the pelvic floor clinical assessment group of the International Continence Society. Neurourol Urodyn 2005;24:374–80.
14. Miller JM, Perucchini D, Carchidi LT, et al. Pelvic floor muscle contraction during a cough and decreased vesical neck mobility. Obstet Gynecol 2001;97:255–60.
15. Constantinou CE, Hvistendahl G, Ryhammer A, et al. Determining the displacement of the pelvic floor and pelvic organs during voluntary contractions using magnetic resonance imaging in younger and older women. BJU Int 2002; 90:408–14.
16. Amarenco G, Ismael SS, Lagauche D, et al. Cough anal reflex: Strict relationship between intravesical pressure and pelvic floor muscle electromyographic activity during cough. Urodynamic and electrophysiological study. J Urol 2005;173:149–52.
17. Deffieux X, Hubeaux K, Porcher R, et al. Pelvic floor muscle activity during coughing: Altered pattern in women with stress urinary incontinence. Urology 2007;70:443–7; discussion 447–8.
18. Deffieux X, Hubeaux K, Porcher R, et al. Abnormal pelvic response to cough in women with stress urinary incontinence. Neurourol Urodyn 2007.
19. Fitzgerald MP, Burgio KL, Borello-France DF, et al. Pelvic-floor strength in women with incontinence as assessed by the brink scale. Phys Ther 2007; 87:1316–24.
20. Black N, Murphy M, Lamping D, et al. Consensus development methods: A review of best practice in creating clinical guidelines. J Health Serv Res Policy 1999;4:236–48.
21. Koch GG, Landis JR, Freeman JL, et al. A general methodology for the analysis of experiments with repeated measurement of categorical data. Biometrics 1977;33:133–58.
22. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543–9.
23. Bo K, Finckenhagen HB. Vaginal palpation of pelvic floor muscle strength: Inter-test reproducibility and comparison between palpation and vaginal squeeze pressure. Acta Obstet Gynecol Scand 2001;80:883–7.
24. Bo K, Lilleas F, Talseth T, et al. Dynamic MRI of the pelvic floor muscles in an upright sitting position. Neurourol Urodyn 2001;20:167–74.
25. Bo K, Finckenhagen HB. Is there any difference in measurement of pelvic floor muscle strength in supine and standing position? Acta Obstet Gynecol Scand 2003;82:1120–4.
26. Frawley HC, Galea MP, Phillips BA, et al. Reliability of pelvic floor muscle strength assessment using different test positions and tools. Neurourol Urodyn 2006;25:236–42.
27. Van Der Velde J, Laan E, Everaerd W. Vaginismus, a component of a general defensive reaction. An investigation of pelvic floor muscle activity during exposure to emotion-inducing film excerpts in women with and without vaginismus. Int Urogynecol J Pelvic Floor Dysfunct 2001;12:328–31.
28. Dietz HP, Shek C. Levator avulsion and grading of pelvic floor muscle strength. Int Urogynecol J Pelvic Floor Dysfunct 2008,19(5):633–6, Epub 2009 Nov 13.
29. Theofrastous JP, Wyman JF, Bump RC, et al. Effects of pelvic floor muscle training on strength and predictors of response in the treatment of urinary incontinence. Neurourol Urodyn 2002;21:486–90.
30. Deffieux X, Hubeaux K, Damphousse M, et al. Perineal neuromuscular fatigue. Ann Readapt Med Phys 2006;49:331–6, 413–7.
31. Theofrastous JP, Wyman JF, Bump RC., et al. Relationship between urethral and vaginal pressures during pelvic muscle contraction. The Continence Program for Women Research Group. Neurourol Urodyn 1997;16:553–8.
32. Peng Q, Jones R, Shishido K, et al. Ultrasound evaluation of dynamic responses of female pelvic floor muscles. Ultrasound Med Biol 2007;33:342–52.
33. Fleiss JL, Cohen J. The equivalence of weighted kappa and the ICC as measures of reliability. Educ Psychol Meas 1973;33:613–9.