ARTICLE

Interrater Reliability of the Observable Movement Quality Scale for Children

Lieke M.A. Dekkers, MSc, PPT;† Maria W.G. Nijhuis-van der Sanden, PhD, PPT;†‡ Marianne Jonker, PhD;‡ Bert J.M. de Swart, PhD;*† Anjo J.W.M. Janssen, PhD, PPT†*

ABSTRACT

Purpose: The authors investigated the interrater reliability, the standard deviation of the random measurement error, and the limits of agreement (LoA) of the Observable Movement Quality (OMQ) scale in children. Movement quality is important in the recognition of motor problems, and the OMQ scale, a questionnaire used by paediatric physical therapists, has been developed for use with an age-specific motor test to observe movement quality and score relative to what is expected for a child's age. ***Method:*** Paediatric physical therapists ($n = 28$; 2 men, 26 women) observed video-recorded assessments of age-related motor tests in children ($n = 9$) aged 6 months to 6 years and filled in the OMQ scale (possible score range 15–75 points). For our analyses, we used linear mixed models without fixed effects. ***Results:*** The interrater reliability was moderate (intra-class correlation coefficient [$ICC_{2,1}$]: 0.67, 95% CI: 0.47, 0.88); neither work setting nor work experience exerted any influence on it. The standard deviation of the random measurement error was 5.7, and the LoA was 31.5. Item agreement was good (proportion of observed agreement [P_o] total 0.82–0.99). ***Conclusion:*** The OMQ scale showed moderate interrater reliability when being used by therapists who were unfamiliar with the questionnaire and who had received only 2 hours of training. Feedback from the participants suggested a need for more comprehensive training in using the OMQ scale in clinical practice.

Key Words: motor skills; movement; paediatrics; reproducibility of results; validation studies.

RÉSUMÉ

Objectif : étudier la fiabilité interévaluateur, l'écart-type (ÉT) de l'erreur de mesure aléatoire et les limites de concordance (LdC) de l'échelle de qualité de mouvements observables (QMO) chez les enfants. La qualité des mouvements est importante pour déceler les problèmes moteurs, et l'échelle de QMO, un questionnaire auquel recourent les physiothérapeutes pédiatriques, a été mise au point pour être utilisée conjointement avec un test de motricité adapté à l'âge afin d'observer la qualité des mouvements et un score relatif aux attentes en fonction de l'âge de l'enfant. ***Méthodologie :*** les physiothérapeutes pédiatriques ($n = 28$; deux hommes, 26 femmes) ont observé des évaluations enregistrées sur vidéo des tests de motricité adaptés à l'âge chez des enfants ($n = 9$) de six mois à six ans et ont rempli l'échelle de QMO (éventail possible des résultats de 15 à 75 points). Dans les analyses, les chercheurs ont utilisé les modèles linéaires mixtes sans effet fixe. ***Résultats :*** la fiabilité interévaluateur était modérée (coefficient de corrélation intraclasse [$CCI_{2,1}$] : 0,67, IC à 95 % : 0,47, 0,88); ni le lieu de travail ni l'expérience de travail n'y exerçait d'influence. L'ÉT de l'erreur de mesure aléatoire était de 5,7, et la LdC, de 31,5. La concordance des points était bonne (proportion du total de concordance observée [P_o] : 0,82 à 0,99). ***Conclusion :*** l'échelle de QMO a révélé une fiabilité interévaluateur modérée lorsqu'elle était utilisée par des physiothérapeutes qui ne connaissaient pas le questionnaire et dont la formation s'était limitée à seulement deux heures. Selon les commentaires de participants, il faudrait une formation plus approfondie pour utiliser l'échelle de QMO en pratique clinique.

Movement quality gives an impression of how movements are controlled and coordinated.[1] Thus, it represents the interaction between personal characteristics and experience, task difficulty, and environmental conditions, and it gives one an insight into the potential of the neurological system to react or adapt to changing conditions.[2] In physical therapy, assessment of movement quality is relevant for recognizing motor problems, evaluating interventions, and predicting recovery.[3–7] To obtain information about movement quality, clinicians and researchers must rely on subjective observation—that is, the process of gathering, organizing, and giving

From the: *Department of Allied Health Studies, HAN University of Applied Sciences; †Donders Institute for Brain Cognition and Behaviour, Department of Rehabilitation, Pediatric Physical Therapy; ‡Radboud Institute for Health Sciences, IQ Healthcare, Radboud University Medical Center, Nijmegen, The Netherlands.

Correspondence to: Lieke M.A. Dekkers, Department of Allied Health Studies, HAN University of Applied Sciences, P.O. Box 6960, 6503 GL Nijmegen, The Netherlands; lieke.dekkers@han.nl.

Contributors: All authors designed the study; or collected, analyzed, or interpreted the data; and drafted or critically revised the article and approved the final draft.

Competing Interests: None declared.

Acknowledgements: The authors thank paediatric physical therapy students Eefje Bongers, Saskia Hock-Roozendaal, and Claudia Holla, who participated in this study as part of their master's of paediatric physical therapy programme.

Physiotherapy Canada **2018; 70(2);113–119; doi:10.3138/ptc.2016-104**

meaning to visual, auditory, and sensory information obtained about a moving person.[2,3]

During the acquisition and re-acquisition of movement, clinicians can observe both quantitative and qualitative changes.[8] Quantitative changes can be seen in people's acquisition of new and more complex motor skills.[9] Currently available discriminative motor tests specifically assess quantitative aspects by comparing individuals with their peers; these tests are norm referenced and validated.

However, changes in the quality of movements demonstrate more subtle characteristics, such as velocity, fluency, accuracy, and automatism of movements.[9] Available and commonly used qualitative measurement instruments focus mostly on specific diagnostic groups, such as children with cerebral palsy (Quality of Function Measure [QFM]),[6] or are designed to assess the functioning of extremities (Quality of Upper Extremity Skills Test [QUEST])[7,10] or children in a specific age range (General Movements [GMs], Infant Motor Profile).[11,12] Earlier studies[2,13] found that descriptions of movement quality are frequently used but not standardized; such descriptions differ among therapists depending on the theoretical construct used in the clinical reasoning, which precludes comparability and longitudinal evaluation.[14,15] Currently, no generic instruments are available to assess movement quality in children over time for all age categories.

To fill this gap, we developed the Observable Movement Quality (OMQ) scale.[13] The OMQ scale is a questionnaire in which each item focuses on an element of observable movement quality (e.g., presence of tremors, fluency, speed of movements). While observing, the therapist is asked to take into account the expected level of performance for a child's age and developmental stage, the task performed, and the environmental circumstances. Therefore, scoring demands an introspective judgment of movement quality based on systematic observations and internal reflection, which incorporates the therapist's knowledge, reasoning, and specific experiences with the target group of children.[2,3] Although the development process established the OMQ scale's content validity,[13] studies on psychometric properties are needed to validate its use in clinical practice.

The aim of this study was to determine the OMQ scale's interrater reliability and standard deviation of the measurement error for paediatric physical therapists who assessed children from ages 6 months to 6 years with different diagnoses. We decided to start with this age group because judging movement quality is more challenging in younger children because of the larger neurobiological changes that occur during early childhood.[11] Moreover, we chose a design using more than two raters to increase its generalizability to clinical practice.[16] The OMQ scale's scoring is, as previously mentioned, based on the introspective judgment of movement quality, which will be influenced by knowledge, reasoning, and personal experiences with the target group of children;[17,18] therefore, we decided to perform two subgroup analyses based on therapists' work setting and years of work experience.

METHODS

This was a cross-sectional reliability study in which paediatric physical therapists judged video recordings of assessments of norm-referenced motor tests of nine children. The medical ethical committee of Radboud University Medical Centre approved the study, which conforms to the principles of the Declaration of Helsinki (registration number 2011/370).

Paediatric physical therapists

This study included a stratified sample of paediatric physical therapists employed in a variety of work settings to guarantee that we included therapists with a variety of clinical expertise: private paediatric physical therapy practices, general hospitals and medical day care centres, and university hospitals and rehabilitation centres. The work settings were located in the southern and central parts of the Netherlands. For each work setting, we included an equal distribution of novice and experienced paediatric physical therapists. The categories of work experience were based on studies by Jensen and colleagues[19,20] and Wainwright and colleagues.[21] To obtain sufficient contrast between novice and experienced physical therapists, we included experienced paediatric physical therapists with 8 years or more years of work experience and novice paediatric physical therapists with 5 or fewer years of work experience. Therapists were verbally informed by the researchers about the study, and those who were interested received an invitation letter in October or November 2011 explaining the study's aim and the total time investment (about 6 h over the course of 5 wk).

The participating therapists signed informed consent forms and received explanations of the privacy rules pertaining to the video recordings of the children. The therapists then received an invitation to a 2-hour training session on scoring the OMQ scale; sessions were organized at nine locations. None of the participants had previous experience with the OMQ scale. The training outlined the purpose of the scale and explained the definitions of the items; all participants received a manual. Participants then watched one video recording of a child with motor problems and filled in the OMQ scale individually. Finally, the scores were compared among the participants; differences and problems in scoring were discussed and unclear issues resolved.

After the training, each therapist received a DVD and numbered OMQ scale scoring sheets for each video recording. The numbers on the scoring sheets corresponded to a unique number for each therapist combined with a number for each child. We asked the therapists to observe the video recording of each of the nine children individually in the order recorded on the DVD

and to score each child's motor quality according to the OMQ scale. The therapists had a maximum of 5 weeks to return the DVDs and OMQ scale scoring sheets to the researchers, using the reply envelope included.

OMQ scale

The OMQ scale[13] was designed for children aged 3 months to 16 years. The 15-item questionnaire needs to be filled in against an age-specific, discriminative motor test to observe and score movement quality relative to what is expected for a child's age. The 15 items are scored on a 5-point Likert scale; thus, total scores range from 15 to 75 (see last presented table for the 15 scale items). Lower scores indicate lower movement quality. Content validity was established during the development of the OMQ scale.[13]

Video-recorded children

For this study, we video recorded nine children; this enabled multiple paediatric physical therapists to observe each child in the same condition. All parents signed informed consent forms for the recording and use of the video for this study.

Eight children were recruited through paediatric physical therapy practices as a representative sample. The inclusion criteria were (1) aged 6 months to 6 years and (2) a diagnosis or indication for treatment by a paediatric physical therapist. We also recruited one typically developing child to ensure that the video recordings included a representation of typical movement quality. We video recorded an age-appropriate motor test during a 1-hour session and used the Alberta Infant Motor Scale to assess children aged 6–13 months;[22] the Bayley Scales of Infant and Toddler Development, Third Edition, to assess children aged 15–23 months;[23] and the Movement Assessment Battery for Children, Second Edition, Dutch version, to assess children aged 3–6 years.[24]

One experienced paediatric physical therapist performed all the motor tests, and another researcher video recorded all the motor tests using a pre-designed protocol. We edited the video recordings to be 15 minutes long per child, ensuring that they showed both fine and gross motor skills and that the aspects of the OMQ scale were observable. The nine video recordings were copied onto a DVD in a random order, using the random number generators menu in IBM SPSS Statistics, version 21 (IBM Corporation, Armonk, NY) to reduce the influence of learning during observation of the nine video cases on the outcome measures.

Statistical methods

We described the characteristics of the therapists and video recorded children to establish the median and range of the continuous variables and the number and percentage of the categorical data. We converted the motor test scores into z scores and calculated OMQ scale total scores as median and range for all therapists and for the two work experience subgroups.

To study the standard deviation of the random measurement error of the OMQ scale, we used a linear mixed model without fixed effects. The dependent variable was the total score on the OMQ scale. Therapists and video-recorded children were treated as random variables (Model A). To study the differences in random measurement error between the two subgroups of paediatric physical therapists (novice, working $\leq$5 y; experienced, working $\geq$8 y) and the three work setting subgroups (paediatric physical therapy practice, general hospitals and medical day care centres, and academic hospitals and rehabilitation centres), we used the same linear mixed model but in a manner (i.e., using a grouping statement in the random intercept statement) that allowed us to estimate a random measurement error per experience group (Model B) and per work setting (Model C).

Initially, we included experience and work setting as independent class variables in Models A, B, and C. However, these terms were always far from statistically significant ($p > 0.80$) and so were omitted from the final models. We calculated the OMQ scale scores obtained from the paediatric physical therapists as a group and by subgroup to obtain the intra-class correlation coefficient type 2:1 ($ICC_{2,1}$), a two-way random effects single-measures model of absolute agreement, standard deviation of the random measurement error, repeatability coefficient (RC), and limits of agreement (LoA). Note that the last two calculations are specific interpretations of the standard deviation of the random measurement error. Furthermore, item agreement is presented as linear-weighted κ, the percentage of observed agreement (P_o), and P_o total, which includes the agreement of a 1-point scoring difference on the Likert scale.

For sample size calculation, we assumed an interrater ICC of 0.8 (i.e., good reliability) and more than 0.6 (i.e., moderate reliability). To obtain a power of 80% ($\alpha = 0.05$, F test), we needed a minimum of 23 observers observing nine different videos.[25] The data were checked for outliers. Statistical analyses were performed in IBM SPSS Statistics and SAS version 9.2 for Windows (SAS Institute, Cary, NC). Two-sided ps < 0.05 were considered statistically significant.

RESULTS

Paediatric physical therapists

Thirty-one paediatric physical therapists agreed to participate in this study. Three female therapists were excluded—one who failed to complete four of nine OMQ scales; a second who misinterpreted the Likert scale and scored inconsistently, as confirmed by outlier analysis; and a third who had technical problems playing the video recordings on the DVD. Table 1 shows the characteristics of the 28 paediatric physical therapists by work setting.

Table 1 Characteristics of Paediatric Physical Therapists by Work Setting

Characteristic	Total ($n = 28$)		Paediatric physical therapy practice		General hospital or medical day care centre		Academic hospital or rehabilitation centre	
	No. (%)	Median (range)	No. (%)	Median (range)	No. (%)	Median (range)	No. (%)	Median (range)
Work experience, y								
≤5	11 (39)	3 (1–5)	5 (42)	2 (1–3)	3 (38)	4 (1–5)	3 (38)	2 (2–4)
≥8	17 (61)	20 (8–29)	7 (58)	19 (10–29)	5 (63)	9 (8–29)	5 (63)	25 (12–28)
Sex								
Male	2 (7)	–	2 (17)	–	0	–	0	–
Female	26 (93)	–	10 (83)	–	8 (100)	–	8 (100)	–

Note: Percentages may not total 100 because of rounding.

Table 2 Characteristics of Video-Recorded Children ($n = 9$) and Results on Motor Test and OMQ Scale

Characteristic			Motor test		OMQ scale total score, median (range)		
Age at video recording, mo	Sex	Diagnosis/indications for treatment	Instrument	*z*-score*	All therapists ($n = 28$)	Work experience, y ≤5 ($n = 11$)	Work experience, y ≥8 ($n = 17$)
13	F	Neuromuscular disorder with hypotonia	AIMS	−7.7	43 (32–51)	46 (35–51)	39 (32–51)
18	M	Trisomy 21	BSID–III	−3.2	51.5 (37–60)	52 (37–55)	50 (38–60)
64	M	Developmental coordination disorder	MABC–2–NL	−3.0	47 (36–60)	51 (44–59)	44 (36–60)
23	F	Trisomy 21	BSID–III	−2.6	53 (32–61)	53 (32–59)	53 (39–61)
14	F	Spastic cerebral palsy, unilateral	BSID–III	−2.2	51 (42–59)	53 (43–59)	50 (42–58)
8	M	Pre-term birth	AIMS	−1.7	65.5 (49–74)	66 (56–74)	64 (49–74)
54	M	Developmental coordination disorder	MABC–2–NL	−1.7	55.5 (47–69)	58 (48–64)	55 (47–69)
6	F	Idiopathic asymmetry†	AIMS	−0.6	67 (48–75)	67 (49–74)	67 (48–75)
38	M	Typical development	MABC–2–NL	1.7	64 (51–74)	62 (51–70)	66 (52–74)

*Standardized score, whereby the raw score is expressed in standard deviation units to compare it with norm scores from typically developing children of the same age (mean = 0; SD = 1).

†Seen in young infants with an asymmetrical head and/or body posture.

OMQ = Observable Movement Quality; F = female; AIMS = Alberta Infant Motor Scale; M = male; BSID–III = Bayley Scales of Infant Development, Third Edition; MABC–2–NL = Movement Assessment Battery for Children, Second Edition, Dutch version.

Of the 28 participants, 26 (93%) women and 2 (7%) men had a median work experience of 11 years (range 1–29 y), and 12 (43%) worked in a paediatric physical therapy practice, 8 (29%) worked in a general hospital or medical day care centre, and 8 (29%) worked in a university hospital or rehabilitation centre. These 28 paediatric physical therapists returned 252 OMQ scale scoring sheets. Median OMQ scale total scores ranged from 43 to 67 for all patients (see Table 2).

Video-recorded children

Table 2 also shows the characteristics of the children—five boys (56%) and four girls (44%), aged 6 months to 5 years, 4 months. As the table shows, diagnoses and indications for treatment by a paediatric physical therapist were common except for the one typically developing child. Motor test *z* scores ranged from –7.7 to 1.7.

Interrater reliability

The interrater reliability was moderate ($ICC_{2,1}$: 0.67; 95% CI: 0.47, 0.88;[26] Table 3). The standard deviation of the random measurement error was 5.7, and no statistically significant differences (i.e., systematic measurement errors) were found among the paediatric physical therapists. The RC was 15.7, representing the value below which the absolute difference between two measurements can be expected only in the presence of random measurement error. The ICC, RC, and LoA across the different subgroups (work experience and work setting) were similar to those for all therapists as a group.

Item agreement

Table 4 shows that the median score for all items on the OMQ scale varied between 3 and 5. For item agreement, κ values for each scale item were low to fair (0.07–0.54), the proportions of observed agreement were fair to good (0.42–0.94), and they improved to good (0.82–0.99) when a 1-point scoring difference on the Likert scale was accepted.

DISCUSSION

In this cross-sectional reliability study, we identified a moderate interrater reliability for the total score on the

Table 3 Interrater Reliability of the OMQ Scale for the Paediatric Physical Therapists by Work Experience and Work Setting

			SD				
Paediatric physical therapists	$ICC_{2,1}$	95% CI	Random measurement error	Between subjects	Among therapists	RC	LoA
All ($n = 28$)*	0.67	0.47, 0.88	5.7	8.0	3.1	15.7	31.5
Work experience†							
≤5 y	0.60	0.37, 0.86	5.8	8.1	2.2	16.1	32.3
≥8 y	0.71	0.51, 0.90	5.6	8.1	3.5	15.4	30.9
Work setting‡							
Paediatric physical therapy practice	0.70	0.49, 0.90	5.4	8.0	2.6	15.1	30.2
General hospital or medical daycare centre	0.62	0.38, 0.87	5.6	8.0	4.0	15.4	30.8
Academic hospital or rehabilitation centre	0.63	0.39, 0.87	6.1	8.0	2.6	16.9	33.9

*Based on Model A.
†Based on Model B.
‡Based on Model C.
OMQ = Observable Movement Quality; $ICC_{2,1}$ = intra-class correlation coefficient, a two-way random effects single-measures model of absolute agreement; RC = repeatability coefficient; LoA = limits of agreement.

Table 4 Agreement of Each Item on the OMQ Scale, Scored by Paediatric Physical Therapists for Video-Recorded Children, Using a 5-Point Likert Scale

			Mean (range)	
Item	Median (range)	Weighted κ* (95% CI)	P_o	P_o total†
1. Appropriate fine motor movements	3 (1–5)	0.35 (0.23, 0.48)	0.52 (0.32–0.61)	0.88 (0.68–1.0)
2. Appropriate gross motor movements	3 (1–5)	0.34 (0.21, 0.46)	0.44 (0.39–0.57)	0.91 (0.75–1.0)
3. Fluency of movements	3 (1–5)	0.24 (0.12, 0.36)	0.45 (0.32–0.57)	0.88 (0.79–0.93)
4. Reduced muscle tone	3 (1–5)	0.54 (0.40, 0.68)	0.58 (0.32–0.82)	0.89 (0.50–1.0)
5. Increased muscle tone	5 (1–5)	0.31 (0.05, 0.57)	0.72 (0.36–0.1)	0.90 (0.68–1.0)
6. Tremors	5 (3–5)	0.07 (–0.07, 0.20)	0.94 (0.86–1.0)	0.99 (0.96–1.0)
7. Slow and/or delayed movements	4 (1–5)	0.42 (0.17, 0.67)	0.52 (0.36–0.89)	0.82 (0.64–1.0)
8. Accelerated and/or abrupt movements	5 (2–5)	0.14 (0.04, 0.24)	0.65 (0.43–1.0)	0.87 (0.75–1.0)
9. Asymmetry in movements	4 (1–5)	0.40 (0.09, 0.72)	0.56 (0.43–0.79)	0.90 (0.89–1.0)
10. Accuracy (well-aimed)	3 (1–5)	0.26 (0.17, 0.36)	0.53 (0.39–0.82)	0.94 (0.82–1.0)
11. Strength regulation	3 (1–5)	0.28 (0.19, 0.37)	0.45 (0.39–0.53)	0.93 (0.82–1.0)
12. Variation in movements	3 (1–5)	0.27 (0.13, 0.42)	0.42 (0.32–0.46)	0.91 (0.82–1.0)
13. Involuntary movements	4 (1–5)	0.19 (0.01, 0.37)	0.53 (0.39–0.85)	0.86 (0.75–0.96)
14. Automated movements	3 (1–5)	0.29 (0.17, 0.41)	0.49 (0.39–0.61)	0.90 (0.82–0.96)
15. Stereotype movements	5 (1–5)	0.31 (0.03, 0.59)	0.71 (0.50–0.92)	0.91 (0.81–1.0)

*Linear weighting.
†Agreement of a 1-point scoring difference on the Likert scale (a score of 1 point higher or lower).
OMQ = Observable Movement Quality; P_o = proportion of observed agreement.

15-item OMQ scale. We found no differences between the two groups of paediatric physical therapists (≤5 or ≥8 y of experience) or among the therapists in the three types of work setting.

This study used video recordings instead of live assessments. Using video recordings both ensured that all therapists observed movement quality under the same circumstances and eliminated the need for multiple observers to examine the children at once. Given that the therapists were observing video recordings, they were unable to interact with the children as they would do in clinical practice, but they did not mention this as a problem. However, they recognized that using video recordings could lead to losing some information for items related to muscle tone.

The lack of difference in interrater reliability based on either work experience or work setting does not support the hypothesis for the expected differences in introspective judgment of movement quality on the basis of clinical experience. One explanation could be that paediatric physical therapists in the Netherlands complete a master's programme in paediatric physical therapy after receiving their bachelor's degree in physical therapy. During this 3-year master's programme, physical therapists work part time with children under the supervision of an experienced colleague—in addition to completing their coursework—to develop clinical expertise by observing, treating, and evaluating interventions. This study focused only on years spent working as a certified paediatric physical therapist and did not include the years

spent working as a general physical therapist. Thus, the differences in outcomes might have been higher if novices with 5 or fewer years of overall working experience had been included.

This study showed a wide range in OMQ scale total scores (17–29 points difference per video-recorded child) and a reasonably large RC (15.7) for the OMQ scale. This could indicate a variation in how the participating therapists interpreted the scoring options for the OMQ scale.[27] For example, as the best-fitting choice for item agreement, we used linear-weighted κ statistics[28]; however, the high level of agreement among the observers led to a low κ value[29,30] for item 6 (0.07), and multiple observers gave the exact same score (i.e., perfect agreement). Furthermore, the small sample size of included children prevented us from performing statistical correlations for the outcomes on the motor tests and OMQ scale. However, ranking the z scores for the motor tests showed that children with higher z scores also showed higher OMQ scale total scores. Only the pre-term infant showed a delay in motor performance (z score –1.7), with a high median OMQ scale total score; this score indicates good quality of movement, which can be observed in pre-term infants at this age,[31] and demonstrates the potential for this child to catch up in motor performance.

Reliability studies are often performed with two or three extensively trained, experienced raters. However, in clinical practice many therapists, both novice and experienced, use a measurement instrument. Reliability studies that use only two or three raters yield results with limited generalizability for the clinical setting.[16] By including 28 paediatric physical therapists employed all over the Netherlands in different work settings and taking into account their years of experience, we increased the generalizability of the results, and we further enhanced them by including children with a variety of diagnoses and who were representative of daily practice.

In this study, the focus was on detecting movement-quality differences in clinical practice rather than on using the OMQ scale for evaluative purposes. In the future, evaluative and longitudinal studies in which intrarater reliability is more relevant will be necessary. The results of this study are motivating and illustrate how training in using the OMQ scale can be improved, including revising the scoring instructions.

Compared with the results of other measurement tools for movement quality,[6,8,10,12] the results of the OMQ scale for interrater reliability were lower. However, both the QFM[6,8] and the QUEST[10] were designed for the cerebral palsy diagnosis group and developed to describe impairment-related movement quality, whereas the OMQ scale was intended to be a generic measurement tool to assess movement quality of the entire body, for all age categories and all diagnoses. In addition, for this study, we developed a 2-hour training session for participating therapists to explain the scale and teach them how to use and interpret it. None of the participating therapists had used the OMQ scale before. In comparison, training in using the GMs,[11] QUEST, and QFM takes 1–2 full days. These factors could have contributed to the lower interrater reliability outcomes in this study.

This study had one limitation: It included two children with Down syndrome for video recording and using a norm-referenced test. Conversations with the therapists revealed that they found it challenging to score these children, possibly because paediatric physical therapists are trained to use the developmental trajectories for such children (as described by Palisano and colleagues[32]) as reference values while observing them. In this study, the therapists had to change perspective and compare their observations with typical development. As the scoring differences on the OMQ scale demonstrated, this perceptual shift proved difficult. During further development of the training for the OMQ scale, we will take these perceived difficulties into account by expanding the focus on observation, regardless of expected motor performance for certain diagnoses or syndromes, supported by videotaped examples.

CONCLUSION

The OMQ scale demonstrates moderate interrater reliability when used by paediatric physical therapists to assess movement quality of children aged 6 months to 6 years. These therapists were unfamiliar with the questionnaire and attended a 2-hour training session on it. Our findings are motivating and indicate that the OMQ scale could be used reliably in clinical practice, although they suggest a need to improve the training. A future study may show that more intensive training can improve the OMQ scale's interrater reliability, a necessary step before determining responsiveness and interpretability. Future clinical cohort studies should also test the effect of the age of a child on interrater reliability and on differences between video and life scoring.

KEY MESSAGES

What is already known on this topic

The assessment of movement quality is relevant for recognizing motor problems, evaluating interventions, and predicting recovery. Currently, no generic instrument is available to assess movement quality over time for all age categories. The Observable Movement Quality (OMQ) scale was developed for this purpose; however, studies on its psychometric properties are needed.

What this study adds

This study demonstrates that it is feasible to rate movement quality using the OMQ scale; however, more comprehensive training is necessary to increase the moderate interrater reliability in therapists unfamiliar with the questionnaire.

REFERENCES

1. Magill R, Anderson D. *Motor learning and control: concepts and applications*. 10th ed. Singapore: McGraw Hill Education; 2014.
2. Skjaerven LH, Kristoffersen K, Gard G. An eye for movement quality: a phenomenological study of movement quality reflecting a group of physiotherapists' understanding of the phenomenon. Physiother Theory Pract. 2008;24(1):13–27. http://dx.doi.org/10.1080/01460860701378042. Medline:18300105
3. Knudson DV, Morrison CS. *Qualitative analysis of human movement*. 2nd ed. Leeds, UK: Human Kinetics; 2002.
4. Heineman KR, Hadders-Algra M. Evaluation of neuromotor function in infancy—A systematic review of available methods. J Dev Behav Pediatr. 2008;29(4):315–23. http://dx.doi.org/10.1097/DBP.0b013e318182a4ea. Medline:18698195
5. Harris SR, Heriza CB. Measuring infant movement: clinical and technological assessment techniques. Phys Ther. 1987;67(12):1877–80. http://dx.doi.org/10.1093/ptj/67.12.1877. Medline:2446339
6. Wright FV, Rosenbaum P, Fehlings D, et al. The Quality Function Measure: reliability and discriminant validity of a new measure of quality of gross motor movement in ambulatory children with cerebral palsy. Dev Med Child Neurol. 2014;56(8):770–8. http://dx.doi.org/10.1111/dmcn.12453. Medline:24702540
7. Sorsdahl AB, Moe-Nilssen R, Kaale HK, et al. Change in basic motor abilities, quality of movement and everyday activities following intensive, goal-directed, activity-focused physiotherapy in a group setting for children with cerebral palsy. BMC Pediatr. 2010;10(1):26. http://dx.doi.org/10.1186/1471-2431-10-26. Medline:20423507
8. Boyce WF, Gowland C, Hardy S, et al. Development of a quality-of-movement measure for children with cerebral palsy. Phys Ther. 1991;71(11):820–8, discussion 828–32. http://dx.doi.org/10.1093/ptj/71.11.820. Medline:1819256
9. Schmidt RA, Wrisberg CA. *Motor learning and performance, a situation-based learning approach*. 4th ed. Leeds, UK: Human Kinetics; 2008.
10. DeMatteo C, Law M, Russell D, et al. The reliability and validity of the Quality of Upper Extremity Skills Test. Phys Occup Ther Pediatr. 1993;13(2):1–18. http://dx.doi.org/10.1080/J006v13n02_01.
11. Hadders-Algra M. General movements: a window for early identification of children at high risk for developmental disorders. J Pediatr. 2004;145(2 Suppl.):S12–8. http://dx.doi.org/10.1016/j.jpeds.2004.05.017. Medline:15292882
12. Hecker E, Baer GD, Stark C, et al. Inter- and intrarater reliability of the Infant Motor Profile in 3- to 18-month-old infants. Pediatr Phys Ther. 2016;28(2):217–22. http://dx.doi.org/10.1097/PEP.0000000000000244. Medline:26901537
13. Janssen AJWM, Diekema ETW, van Dolder R, et al. Development of a movement quality measurement tool for children. Phys Ther. 2012;92(4):574–94. http://dx.doi.org/10.2522/ptj.20100354. Medline:22173795
14. Edwards I, Jones M, Carr J, et al. Clinical reasoning strategies in physical therapy. Phys Ther. 2004;84(4):312–30, discussion 331–5. Medline:15049726.
15. Wainwright SF, Shepard KF, Harman LB, et al. Novice and experienced physical therapist clinicians: a comparison of how reflection is used to inform the clinical decision-making process. Phys Ther. 2010;90(1):75–88. http://dx.doi.org/10.2522/ptj.20090077. Medline:19926680
16. Van Genderen FR, de Bie RA, Helders PJM, et al. Reliability research: towards a more clinically relevant approach. Phys Ther Rev. 2003;8(4):169–76. http://dx.doi.org/10.1179/108331903225003172.
17. Hayward LM, Black LL, Mostrom E, et al. The first two years of practice: a longitudinal perspective on the learning and professional development of promising novice physical therapists. Phys Ther. 2013;93(3):369–83. http://dx.doi.org/10.2522/ptj.20120214. Medline:23064736
18. Black LL, Jensen GM, Mostrom E, et al. The first year of practice: an investigation of the professional learning and development of promising novice physical therapists. Phys Ther. 2010;90(12):1758–73. http://dx.doi.org/10.2522/ptj.20100078. Medline:20930050
19. Jensen GM, Shepard KF, Hack LM. The novice versus the experienced clinician: insights into the work of the physical therapist. Phys Ther. 1990;70(5):314–23. http://dx.doi.org/10.1093/ptj/70.5.314. Medline:2333330
20. Jensen GM, Gwyer J, Hack LM, et al. Expertise in physical therapy practice. 2nd ed. St. Louis, MO: Saunders Elsevier; 2007.
21. Wainwright SF, Shepard KF, Harman LB, et al. Factors that influence the clinical decision making of novice and experienced physical therapists. Phys Ther. 2011;91(1):87–101. http://dx.doi.org/10.2522/ptj.20100161. Medline:21127167
22. Piper MC, Darrah J. Motor assessment of the developing infant. Vol. 81, 1st ed. Philadelphia: W.B. Saunders; 1994.
23. Bayley N. *Bayley Scales of Infant and Toddler Development, 3rd edition*. San Antonio (TX): Psychological Cooperation; 2006.
24. Henderson SE, Sugden DA, Barnett LA. *Movement Assessment Battery for Children, second edition, manual, Nederlandse bewerking*. London: Pearson Education International; 2010.
25. Shoukri MM, Asyali MH, Donner A. Sample size requirements for the design of reliability study: review and new results. Stat Methods Med Res. 2004;13(4):251–71.
26. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42. http://dx.doi.org/10.1016/j.jclinepi.2006.03.012. Medline:17161752
27. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol. 2008;31(4):466–75. http://dx.doi.org/10.1002/uog.5256. Medline:18306169
28. Schouten HJA. Measuring pairwise interobserver agreement when all subjects are judged by the same observers. Stat Neerl. 1982;36(2):45–61. http://dx.doi.org/10.1111/j.1467-9574.1982.tb00774.x.
29. de Vet HCW, Mokkink LB, Terwee CB, et al. Clinicians are right not to like Cohen's κ. BMJ. 2013;346:f2125. http://dx.doi.org/10.1136/bmj.f2125. Medline:23585065
30. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43(6):543–9. http://dx.doi.org/10.1016/0895-4356(90)90158-L. Medline:2348207
31. van Haastert IC, de Vries LS, Helders PJ, et al. Early gross motor development of preterm infants according to the Alberta Infant Motor Scale. J Pediatr. 2006;149(5):617–22. http://dx.doi.org/10.1016/j.jpeds.2006.07.025. Medline:17095330
32. Palisano RJ, Walter SD, Russell DJ, et al. Gross motor function of children with Down syndrome: creation of motor growth curves. Arch Phys Med Rehabil. 2001;82(4):494–500. http://dx.doi.org/10.1053/apmr.2001.21956. Medline:11295010