



HAL
open science

Rasch analysis of the motor function measure in patients with congenital muscle dystrophy and congenital myopathy

Carole Vuillerot, P. Rippert, V. Kinet, A. Renders, M. Jain, M. Waite, A. M. Glanzman, F. Girardot, D. Hamroun, J. Iwaz, et al.

► To cite this version:

Carole Vuillerot, P. Rippert, V. Kinet, A. Renders, M. Jain, et al.. Rasch analysis of the motor function measure in patients with congenital muscle dystrophy and congenital myopathy. Archives of Physical Medicine and Rehabilitation, 2014, 95 (11), pp.2086-95. 10.1016/j.apmr.2014.06.005 . hal-02045540

HAL Id: hal-02045540

<https://univ-lyon1.hal.science/hal-02045540v1>

Submitted on 21 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Published in final edited form as:

Arch Phys Med Rehabil. 2014 November ; 95(11): 2086–2095. doi:10.1016/j.apmr.2014.06.005.

Rasch Analysis of the Motor Function Measure in Patients With Congenital Muscle Dystrophy and Congenital Myopathy

Carole Vuillerot, MD, PhD^{a,b,c,d,e}, Pascal Rippert, PhD^f, Virginie Kinet, PT^g, Anne Renders, MD^g, Minal Jain, PT, DSc, PCS^h, Melissa Waite, PT^h, Allan M. Glanzman, DPT, PCSⁱ, Francoise Girardot, PT^a, Dalil Hamroun, MD^j, Jean Iwaz, PhD^{b,c,d,k}, René Ecochard, MD, PhD^{b,c,d,k}, Susana Quijano-Roy, MD, PhD^{l,m}, Carole Bérard, MD^a, Isabelle Poirot, MD, PhD^a, Carsten G. Bönnemann, MD^e, and the CDM MFM Study Group

^aHospices Civils de Lyon, Hôpital Femme-Mère-Enfant, L'Escale, Service de Médecine Physique et de Réadaptation Pédiatrique, Bron, France

^bUniversité de Lyon, Lyon, France

^cUniversité Lyon I, Villeurbanne, France

^dCNRS UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Biostatistique Santé, Pierre-Bénite, France

^eNational Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD

^fHospices Civils de Lyon, Pôle Information Médicale, Évaluation, Recherche, Lyon, France

^gCliniques Universitaires Saint-Luc, Centre de Référence des maladies neuromusculaires, Université Catholique de Louvain, Brussels, Belgium

^hClinical Research Center, National Institutes of Health, Bethesda, MD

ⁱPhysical Therapy Department, The Children's Hospital of Philadelphia, Philadelphia, PA

^jCentre Hospitalo-Universitaire de Montpellier, Hôpital Arnaud de Villeneuve, Montpellier, France

^kHospices Civils de Lyon, Service de Biostatistique, Lyon, France

^lAP-HP, Hôpitaux Universitaires Paris-Ile-de-France Ouest, Pôle pédiatrique, Centre de Référence Maladies Neuromusculaires (GNMH), Hôpital Raymond Poincaré, Garches, France

^mUniversité Versailles Saint-Quentin, St Quentin, France

Abstract

Corresponding authors: Carole Vuillerot, MD, PhD, Porter Neuroscience Research Center, National Institute of Neurological Disorders and Stroke/National Institutes of Health, 10 Center Dr, Bldg 10, Rm 12N210, Bethesda, MD 20892. carole.vuillerot@gmail.com. Carsten G. Bönnemann, MD, Neuromuscular and Neurogenetic Disorders of Childhood Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD. carsten.bonnemann@nih.gov.

Disclosures: none.

Objectives—To monitor treatment effects in patients with congenital myopathies and congenital muscular dystrophies, valid outcome measures are necessary. The Motor Function Measure (MFM) was examined for robustness, and changes are proposed for better adequacy.

Design—Observational study based on data previously collected from several cohorts.

Setting—Nineteen departments of physical medicine or neuromuscular consultation in France, Belgium, and the United States.

Participants—Patients (N=289) aged 5 to 77 years.

Interventions—None.

Main Outcome Measures—A Rasch analysis examined the robustness of the MFM across the disease spectrum. The 3 domains of the scale (standing position and transfers, axial and proximal motor function, and distal motor function) were independently examined with a partial credit model.

Results—The original 32-item MFM did not sufficiently fit the Rasch model expectations in either of its domains. Switching from a 4- to a 3-category response scale in 18 items restored response order in 16. Various additional checks suggested the removal of 7 items. The resulting Rasch-scaled Motor Function Measure with 25 items for congenital disorders of the muscle (Rs-MFM25^{CDM}) demonstrated a good fit to the Rasch model. Domain 1 was well targeted to the whole severity spectrum—close mean locations for items and persons (0 vs 0.316)—whereas domains 2 and 3 were better targeted to severe cases. The reliability coefficients of the Rs-MFM25^{CDM} suggested sufficient ability for each summed score to distinguish between patient groups (0.9, 0.8, and 0.7 for domains 1, 2, and 3, respectively). A sufficient agreement was found between results of the Rasch analysis and physical therapists' opinions.

Conclusions—The Rs-MFM25^{CDM} can be considered a clinically relevant linear scale in each of its 3 domains and may be soon reliably used for assessment in congenital disorders of the muscle.

Keywords

Congenital muscular dystrophy; Congenital myopathy; Disability evaluation; Outcome measures; Rehabilitation

Congenital myopathies (CMs) and congenital muscular dystrophies (CMDs) form a group of early-onset muscle disorders with a large spectrum of phenotypes (progressive or stable courses, ambulant or nonambulant patients, with or without respiratory insufficiency, with or without cognitive impairment). However, significant early weakness and multiple joint contractures are very common; they challenge motor outcome scales¹ and may lead to inaccurate motor-function assessments. It is therefore essential that a scale destined to measure motor-function changes in this population provide clinically meaningful and scientifically robust data, especially for use in clinical trials.²⁻⁴

In 2011, we reviewed many rating scales for their adequacy to CMDs. The Motor Function Measure (MFM) showed encouraging results in a sample of 52 patients.¹ However, 1 major limitation was that prominent upper and lower extremity contractures interfered with the

patients' ability to perform some items. Because the MFM cannot be assumed to fit all types of neuromuscular disorders equally well (ie, all items useful, achievable, and able to cover the whole severity spectrum), its applicability and validity had to be reviewed in a larger sample of patients in an attempt to identify inadequate or suboptimal items.

Within this context, the Rasch analysis, developed by Georg Rasch,⁵ is a rigorous approach for an in-depth understanding of a scale's metrological properties⁶⁻⁹ and an evaluation of properties not analyzed by the classical test theory—for example, how well an item performs in terms of relevance or usefulness in measuring the underlying test construct, the amount of the construct targeted by each question, the possible redundancy of an item regarding other items in the scale, and the appropriateness of response categories along the entire range of disease severity.¹⁰ The Rasch model assumes that the probability of a participant's endorsing (ie, ability to perform) an item is a logistic function of the relative difference between the item's location (its difficulty) and the person's location (his or her ability). In other words, for 2 patients with different abilities, the more able patient should have a higher probability of achieving the task (or item), and for 2 items with different difficulties, the more difficult item should have a lower probability of success than the other, regardless of the patients' ability.

Using statistical and clinical approaches, a rigorous Rasch-based evaluation of the MFM was thus performed in 289 patients with CM or CMD. Previous results have shown the validity of the MFM using classical test theory methodology^{11,12}; the present study examined whether the MFM fulfills all Rasch model expectations and proposes changes to improve the scale's internal validity.

Methods

Ethics

Ethical approvals were obtained from the Comité de Protection des Personnes Lyon Sud Est II and from the institutional review boards of the National Institutes of Health and the Children's Hospital of Philadelphia. The participants' consents were obtained according to the Declaration of Helsinki.

Data collection

All the participants had a diagnosis of CMD or CM (clinical, muscle biopsy \pm genetic testing).

In this study, single scores were used; the first score was used only when more than 1 score was collected. MFM scores were compiled from 4 independent MFM databases in France, the United States, and Belgium. The French secure Web-based database collects MFM data on patients with neuromuscular diseases from a network of 17 specialized centers in France. A request to use the data was approved by the MFM committee (see www.mfm-nmd.org). The U.S. data were drawn from a National Institute of Health protocol titled "Clinical and Molecular Manifestations of Neuromuscular and Neurogenetic Disorders of Childhood" (Clinical Trial Registration No.: NCT01568658) and from The Children's Hospital of

Philadelphia Natural History Study. Additional data were collected from the Center for Neuromuscular Diseases, Hôpital St Luc, Brussels, Belgium.

Clinical assessment

The MFM-32 is a physician- or physical therapist-administered physical test that consists of 32 items (or tasks) in 3 functional domains: D1, standing and transfers; D2, axial and proximal motor function; and D3, distal motor function.¹³ Its scoring system uses a 4-point Likert scale based on the increasing ability to perform each task without assistance: 0, cannot initiate the task or maintain the starting position; 1, partially completed the task; 2, completed the task with compensations, slowness, or obvious clumsiness; and 3, completed the task “normally,” the standard pattern (see the details on scoring each item in the User’s Manual available at www.mfm-nmd.org). The MFM is applicable with the same reliability to children and adolescents.¹³

In using the MFM in patients with CM or CMD, some physical therapists had informally said that some items are poorly adapted to patients with severe elbow, neck, or finger contractures that prevented them from having the correct starting positions. The current MFM User’s Manual recommends scoring these items 0. To identify these problems more formally, all the items were examined by 6 expert neuromuscular physical therapists familiar with the MFM in patients with CM and CMD, who had to identify the items they would change or remove to increase the applicability of the MFM in these patients. The physical therapists were also asked to rank the items from 0 to 5 in the increasing order of difficulty.

Statistical analysis

Descriptive statistics were used to summarize the patients’ characteristics and the MFM scores. The means, SD, and ranges of the MFM scores (D1, D2, D3, and total) were analyzed by CMD and CM disease. For each item, the median and range of item difficulty and the percentage of physical therapist agreement to change or remove the item were also calculated.

The Rasch evaluation was performed using RUMM2030 software.^{14, a} A preliminary Rasch analysis on all 32 items having shown a severe misfit due to multidimensionality, distinct analyses were performed for the 3 MFM domains (D1, D2, and D3). The partial-credit model (which allows each item to have its own threshold parameters) was used with each domain because the likelihood ratio test was significant ($P < .0001$). This indicated that the rating scale model (which requires equivalent thresholds across all items) was not appropriate⁷ (appendix 1).

The Rasch analysis followed the procedures and guidelines recommended by several authors^{7, 15–19} (see appendix 1 and fig 1). The summary item-trait interaction statistics reflect the fit of the observed data to the model’s expectations and is represented by the chi-square test. A significant chi-square test ($P < .05$) indicates that the data do not fit the model; the reasons for the misfit should then be explored. For instance, when there is too little

^aRUMM2030, version 5.1. [computer software]; RUMM Laboratory, Perth, Western Australia. Available at: <http://www.rummlab.com.au/>.

discrimination between 2 response categories on an item, a given patient may have a higher probability of scoring lower than would a less able patient (disordered categories concept). A disordered category can be identified with category probability curves that express, for each item, the probability of a given category (0–3) as a function of person ability. In such a case, collapsing the uncertain categories into a single category can improve the scale fit and restore the ordered categories. When different groups within the sample (eg, men vs women, children vs adults, and CM vs CMD) respond differently to an item despite equal levels of the underlying characteristic being measured (differential item functioning [DIF]), a deletion of the item can be considered. When the response to a given item is influenced by the response to a previous one (location dependence), both items have to be reentered into the model to correct the misfit. The targeting of the scale to the study population and the assumption of unidimensionality in each domain were checked. Removing items was considered in case of persistent misfit after the previous steps. Different criteria were used to identify the items to be removed: fit residuals more than 2.5 or less than -2.5 or exceeding the Bonferroni-adjusted alpha, ceiling effect, and the percentage of physical therapists agreeing on the removal, and how well, for each item, the observed model tended to fit the expected model in groups of responders across the trait (all items were viewed graphically). An iterative procedure was used: 1 item was removed at a time, beginning with the item with the highest number of removal criteria, and the fit reestimated until nonsignificance of the chi-square test.

The Person separation index is provided by RUMM 2030 as an estimation of the internal consistency of the scale and the power of the scale to discriminate among respondents with different levels of motor function. The Person separation index is interpreted similarly to Cronbach's α : 0.7 is the minimal acceptable value for group or research use and .80 for individual or clinical use.^{7,20}

Results

Descriptive analysis results

The analysis was conducted on data on 289 patients aged 5 to 77.2 years. Table 1 shows the descriptive statistics. The most affected patients were those with laminin alpha2-deficiency; 52% of them were nonambulant. This disease showed the lowest mean MFM scores: 21.3% \pm 29.9% for D1, 62.1% \pm 32.9% for D2, 71.1% \pm 26.2% for D3, and 49.7% \pm 28.3% for the total score (see table 1).

Figure 2 shows the MFM total scores by age and CMD diagnosis. Overall, the scores of patients with CM and CMD demonstrated similar scatter patterns, indicating homogeneous distributions of patients across the entire severity spectrum; however, patients with CMD appeared to be more impaired than patients with CM (MFM total scores, 60.8 \pm 24.6 vs 73.5 \pm 21.8, respectively) (independent-sample *t* test; *P* = .026).

Table 2 shows the median and ranges of ranks given by the physical therapists to item difficulties. Unanimously, items 30 (Run), 31 (Hop on 1 foot), and 32 (Squat and stand up) were labeled as the most difficult and items 1 (Turn the head), 12 (Sit on a chair), 17 (Pick

up the coins), and 22 (Place the finger on each of the drawings) as the easiest. Overall, D1 items were considered more difficult than D2 or D3 items.

Rasch analysis–driven changes to the MFM

The fit to the Rasch model was insufficient for the original MFM D1 domain (13 items) (table 3). Five items showed disordered categories and were thus rescored with 3-response categories (0, 1, 2) as suggested by the items' category probability curves. Statistically significant DIFs were observed for age and sex for item 6 (Raise the pelvis) (mean squares, 31.8 and 14.6, respectively; $F=30.8$ and 13.1 , respectively, with $df=1$; $P<.0001$ and $.0004$, respectively), suggesting its deletion. Items 25 (Stand straight without support) and 8 (Sit up from supine position) were also deleted because they had the highest number of deletion criteria (see the deletion criteria in table 2). The fit to the expectations of the Rasch model was reached only after the removal of both items. There was no evidence of location dependence. A high Person separation index (.89) was obtained for the Rasch-scaled 10-item MFM D1 domain. After the deletions, items 31 (Hop on 1 foot) and 32 (Squat and stand up) were identified as the most difficult (item locations, 1.353 and 1.417, respectively), whereas item 12 (Sit on a chair) was identified as the easiest.

The original MFM D2 domain did not sufficiently fit the Rasch model expectations (see table 3). Rescoring 9 items with disordered categories failed to restore the ordered categories for items 9 (Maintain the seated position) and 10 (Lean forward) because the data remained dichotomous: these items were deleted. Statistically significant DIFs were observed for age for item 7 (Roll onto the stomach) (mean square, 7.47; $F=14.34$ with $df=1$; $P=.0002$), suggesting its deletion. All other items were kept because, according to the deletion criteria, there was no consensus on other problematic items and because the deletion of other items did not correct the misfit. A location dependence (correlation >0.3) was found between items 2 (Raise the head) and 16 (Reach for a pencil) and between items 3 (Flex the hip) and 15 (Place both hands on the head). Combining these 2 pairs of items and reentering them in the analysis improved the statistical parameters and solved the misfit (see table 3). A Person separation index of .77 was obtained with the Rasch-scaled 9-item MFM D2 domain. Item 2 (Raise the head) was identified as the most difficult, whereas item 23 (Place both hands on the table) was identified as the easiest (item locations, 1.907 and -1.125 , respectively; see table 2).

The original MFM D3 domain did not sufficiently fit the Rasch model expectations (see table 2). Four items with disordered categories were rescored with 3 response categories after collapsing categories 1 and 2. There were no cases of DIF by age, sex, or diagnosis, suggesting that these factors were not the causes of the misfit. The removal criteria identified items 4 (Pull up the foot) and 22 (Place a finger on each of the drawings) as candidates for removal (see table 3), but only item 4 removal improved the fit statistic. Location dependence (correlation $>.3$) was found between items 19 (Pick up and draw loops) and 21 (Pick up and turn a ball in the air): they were therefore combined. Reentering the new item with the others in the analysis improved the statistical parameters and solved the misfit (fig 3). The final analysis of the Rasch-scaled 6-item MFM D3 domain showed a fit to the expectations of the Rasch model (see table 3) with a Person separation index of .66

(Cronbach $\alpha=0.78$). Item 20 (Tear a 4-folded sheet of paper) was identified as the most difficult, and item 22 (Place the finger on each of the drawings) as the easiest (item locations, 1.524 and -1.319 , respectively; see table 2).

Item targeting

The item location ranged from -1.416 to 1.416 for D1, from -1.131 to 1.908 for D2, and from -1.299 to 1.172 for D3, indicating that the Rasch-scaled MFM achieved a good continuum with little overlap and very few items per domain measuring the same ability level (see fig 3).

Figure 3 shows an adequate correspondence between the distributions of person measurements and item locations (upper vs lower histogram) for D1; the mean person location logit value (0.371 ± 1.722) was close to the mean of item locations (0.000 ± 1.115). For D2 and D3, the distributions of person measurements did not correspond to the distributions of item locations; most of the persons were located on the right side of the histogram, and the mean person location logit values were not close to 0 (1.632 ± 1.496 for D2 and 2.358 ± 1.621 for D3). These positive values suggest a ceiling effect for D2 and D3 owing to the lack of more difficult items.

Unidimensionality

For each domain, the assumption of unidimensionality was checked. Adequate results were obtained for D1, D2, and D3 (5.9%, 4.2%, and 1.0%, respectively, the expected value being $<5\%$).

Discussion

Reliably capturing a significant change in patient status due to a new treatment or intervention requires complete confidence in the outcome-measuring tool. Here, Rasch analysis was used to review the metrological properties of the MFM in patients with CMD and CM. This disease group challenges the application of established motor testing scales because of the early onset of significant weakness (inability to walk and contractures) in many patients. The distributions of the raw MFM total scores according to age were similar in patients with either CM or CMD; this indicated similar functional profiles and allowed the joint study of these patients. The Rasch-scaled Motor Function Measure with 25 items for congenital disorders of the muscle (Rs-MFM25^{CDM}), which would include the above-mentioned changes and apply to patients with congenital disorders of the muscle, demonstrated an adequate fit to the Rasch model with acceptable reliability, effective targeting of the 3 domains, and no evidence of DIF or failure of items to fit a latent trait of motor function. This was supported by a large sample size (289 patients) that exceeded the requirement for a stable model.¹⁷

In health-related research, a Rasch analysis is often used when a set of questions or administered items are intended to be summed together to provide a total score.⁷ In the field of neuromuscular diseases, a Rasch analysis has been successfully used to develop ACTIVLIM, a self-report measurement of activity limitation, by selecting items that fit the Rasch model expectations (lack of DIF and high unidimensionality).²¹

A Rasch analysis has also been used to review and improve the metrological properties of already existing ordinal scales. Mayhew et al²² supported that the North Star Ambulatory Assessment is a reliable measure of ambulatory function in patients with Duchenne muscular dystrophy and reported an adequate targeting and a high Person separation index (=0.91). Their Rasch analysis showed no disordered categories, probably because 3-response category items are less ambiguous and source of misfit than 4-response category items; however, less response categories may decrease the ability of a scale to measure score changes.²³

A Rasch analysis is especially needed to enhance the sensitivity to change of a scale in a specific group of patients and allow its use in clinical trials.²⁴ This kind of analysis was recently used by Cano et al⁹ for spinal muscular atrophy outcome measures. It is used here to identify some items that affect the validity of MFM-32 and introduce the minimal changes that can ensure the fit of the data to the Rasch model.

First, 7 items (items 4, 6–10, and 25) had to be removed on predefined criteria.¹⁹ A qualitative analysis has shown that scoring of some of them was potentially influenced by a failure to take up the starting position (items 9 and 10). Joint limitations or poor alignment could also prevent reaching the maximal score (items 6, 7, and 25). Item 4 was misfit, probably because it is the only D3 item related to the distal motor function of the lower limb whereas all others are related to the upper limb.

Second, after item deletion, only 14 items identified with disordered categories were rescored with 3 response categories because rescored all 25 items might have lowered the ability of the MFM to capture changes.²³

To ensure the clinical practical relevance of the MFM scale during the improvement process, physical therapists' opinions on MFM-32 items' relevance and difficulty were compared with the results of the Rasch analysis. The 2 rankings of item "difficulty" agreed almost perfectly. Items 31 (hopping) and 32 (squatting), which require advanced motor functions, were identified as very difficult by both rankings. In domain D2, items 2 (Flexion of the head) and 14 (Extension of the head) were identified as the most difficult by both rankings; this is due to the occurrence of significant neck flexor and extensor weakness and/or contractures in patients with CM and CMD. In domain D3, item 22 (Place the finger in 8 drawings) was unanimously identified by the physical therapists as the easiest item and recommended for removal because of its low discriminant value. The Rasch analysis also identified item 22 as easy but recommended to keep it because it showed a sufficient discriminant value, which confirmed our clinical impression. Item 20 (Tear a 4-folded sheet of paper) was also identified by both rankings as the most difficult D3 item; significant contractures of the fingers associated with weakness occur in CM and CMD, leading to reduced distal hand motor function.

Scale-to-sample targeting of the Rasch-scaled MFM D1 was good; however, an inadequate targeting was found for domains D2 and D3, suggesting that the Rasch-scaled MFM D2 and D3 would better target the more severely affected patients. This was considered acceptable

because the less severely affected patients (ie, those with the highest level of motor function) can be adequately evaluated by domain D1 items.

With regard to the reliability of the Rs-MFM25^{CDM}, we found the internal consistency, as measured by the Person separation index, to be excellent for D1 (Person separation index=.90), good for D2 (Person separation index=.77), and acceptable for D3 (Person separation index=.67). Thus, the sum score of each of the 3 domains has sufficient reliability to distinguish between patient groups for research purposes.²⁵ The 3 distinct Rasch analyses for the 3 distinct domains of the MFM allowed the construction of a summed score for D1, D2, and D3 but not a total score because of the assumption of unidimensionality.

Study limitations

First, the value of the results depends obviously on the representativeness of the study population. Here, despite the rarity of the disease, the sample size may be considered as fairly representative of the population of subjects with congenital disorders of the muscle.

It is worth noting here that the present study undertook a Rasch analysis of the MFM-32 and did not concentrate on the exact way of rewording the items that would replace item pairs with location dependence, on defining new response options, and on the way of obtaining the final score. In other words, though ready for research work, the present Rs-MFM25^{CDM} is not definitive and not ready for everyday clinical use. However, these steps are under way.

In addition, because of item deletion, the Rs-MFM25^{CDM} should undergo a separate assessment of its validity by another team in another context and should be retested in another cohort of patients with CM and/or CMD to compare its validity and sensitivity to change versus the original MFM-32. The ongoing work on this topic is promising. Once confirmed, the results will lead to shortening the number of items of the MFM, the overall duration of the tests, and the fatigue of the patients, especially the youngest ones. Furthermore, we may derive short versions specific to patients with CMD or CM whose scores, reproducibly converted into Rasch measurements, will be linear and will allow longitudinal follow-ups and comparisons between patients.

Conclusions

The Rs-MFM25^{CDM} can be assumed as a linear scale in each of its 3 domains. Its practical clinical relevance is confirmed by the agreement between the Rasch analysis results and the physical therapists' opinions concerning item difficulty. These findings promote the Rs-MFM25^{CDM} as an outcome measure suitable for interventional clinical trials in patients with both CM and CMD.

Acknowledgments

Supported by intramural funds of the National Institute of Neurological Disorders and Stroke, the Association Française contre les Myopathies, the Société Française de Médecine Physique et de Réadaptation, the Société Francophone d'Etude et de Recherche sur les Handicaps de l'Enfance, the Philippe Foundation, and Bouillat Terrier and Deage.

We thank Pierre Bigot, MD, for his comments and suggestions on this work. We thank the members of the MFM CDM Group who performed the MFM tests. The group is composed of the following French, Swiss, Belgian, and

American professionals involved in the validation of the MFM in congenital disorders of the muscle: A. Renders, MD, D. Laridant, MD, and V. Kinet, PT (Cliniques Universitaires St Luc, Brussels); V. Sperhrs-Ciaffi, MD (Pediatric Center for Neuromuscular Disease, CHUV, Lausanne); J. Datsgir, MD, K. Meilleur, PhD, S. Donkervoort, MS, CGC, M. Leach, MSN, PNP-BC, and A. Rutkowski, MD (CMD COM core study group); M. Fournier-Méhouas, MD, and V. Tanant, PT (Hôpital de l'Archet, Nice); H. Rauscent, MD, and F. Letanoux, PT (CHU Rennes Pontchaillou); S. Ragot, MD, I. Mugnier, MD, and C. Capello, PT (Hôpital Brabois, Nancy); J.P. Vadot, MD, C. de Lattre, MD, F. Margirier, PT, D. Vincent-Genod, PT, A. Berruyer, PT, A. Barrière, OT, F. Bouhour, MD, and J.F. Remec, PT (CHU Lyon, Hopital Femme-Mère-enfant); J. Lachanat, MD, and D. Denis, PT, (Fondation Richard, Lyon); J.Y. Mahé, MD, and C. Nogues, PT (Centre de Pen Bron, La Turballe); S. Chabrier, MD, C. Gayet, PT, M.C. D'Anjou, MD, L. Feasson, MD, PhD, and A. Jouve, PT (Hôpital Bellevue, Saint Etienne); J.A. Urtizbera, MD, PhD, and A. Cobo, PT (Hendaye); C. Themar Noel, MD, T. Stojkovic, MD, C. Dejoux, PT (Institut de Myologie, Hôpital Pitié-Salpêtrière, Paris); S. Quijano, MD, PhD, N. Pelligrini, MD, N. Vedrenne, PT, and S. Morel-Lelu, PT (Hôpital Raymond Poincaré, Garches); M. Marpeau, MD, F. Barthel, MD, D. Traubaud, PT, and M. Vercaemer, PT (Centre St Jean de Dieu, Paris); V. Tiffereau, MD, PhD, E. Hovart, PT (Hôpital Swinghedaw, Lille); A. Carpentier, MD, and I. Bourdeauducq, PT (Centre Marc Sautet, Villeneuve d'Ascq); A. Labarre-Villa, MD, M.C. Commare, MD, M. Mahot, PT, and V. Farigoule, PT (Centre hospitalier de Grenoble); M.A. Barthez, MD, and C. Destremaut, PT (CHU Clocheville, Hôpital des enfants), S. Pellieux, MD, and C. Pannet, PT (CHRU Trousseau); J.Y. Salle, MD, and D. Varnoux, PT (CHU Dupuytren), P. Kiény, MD, and G. Morel, PT (Résidence la Forêt - Centre AFM); D. Pichancourt, MD, and N. Vedrenne, PT (CHR Pierre le Damany); and M. Campech, MD, and F. Robert, PT (CHR Félix Guyon-la Réunion).

List of Abbreviations

CM	congenital myopathy
CMD	congenital muscular dystrophy
DIF	differential item functioning
MFM	Motor Function Measure
Rs-MFM25^{CDM}	Rasch-scaled MFM with 25 items for congenital disorders of the muscle

References

1. Bonnemann CG, Rutkowski A, Mercuri E, Muntoni F. 173rd ENMC International Workshop: congenital muscular dystrophy outcome measures 5–7 March 2010, Naarden, The Netherlands. *Neuromuscul Disord.* 2011; 21:513–22. [PubMed: 21641800]
2. Fermanian J. Validation of assessment scales in physical medicine and rehabilitation: how are psychometric properties determined? *Ann Readapt Med Phys.* 2005; 48:281–7. [French]. [PubMed: 15923054]
3. Hobart JC, Cano SJ, Zajick JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol.* 2007; 6:1094–105. [PubMed: 18031706]
4. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess.* 2009; 13:iii, ix–x, 1–177.
5. Rasch, G. Probabilistic models for some intelligence and attainment tests. Texas: Univ of Chicago Pr; 1981. p. 199
6. Conrad KJ, Smith EV Jr. International conference on objective measurement: applications of Rasch analysis in health care. *Med Care.* 2004; 42:11–6. [PubMed: 14707750]
7. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum.* 2007; 57:1358–62. [PubMed: 18050173]
8. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. *Int J Nurs Stud.* 2009; 46:380–93. [PubMed: 19059593]

9. Cano SJ, Mayhew A, Glanzman AM, et al. Rasch analysis of clinical outcome measures in spinal muscular atrophy. *Muscle Nerve*. 2014; 49:422–30. [PubMed: 23836324]
10. Reeve, BB.; Fayers, P. Applying item response theory modelling for evaluating questionnaire item and scale properties. In: Fayers, P.; Hays, R., editors. *Assessing quality of life in clinical trial: methods and practice*. Oxford: Oxford Univ Pr; 2005. p. 55-74.
11. Vuillerot C, Payan C, Girardot F, et al. Responsiveness of the motor function measure in neuromuscular diseases. *Arch Phys Med Rehabil*. 2012; 93:2251–2256. e1. [PubMed: 22705238]
12. Vuillerot C, Payan C, Iwaz J, Ecochard R, Berard C. Responsiveness of the motor function measure in patients with spinal muscular atrophy. *Arch Phys Med Rehabil*. 2013; 94:1555–61. [PubMed: 23380348]
13. Berard C, Payan C, Hodgkinson I, Fermanian J. A motor function measure for neuromuscular diseases: construction and validation study. *Neuromuscul Disord*. 2005; 15:463–70. [PubMed: 16106528]
14. Andrich, D.; Lyne, A.; Sheridan, B.; Luo, G. RUMM 2030. Perth: RUMM Laboratory; 2010.
15. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol*. 2007; 46:1–18. [PubMed: 17472198]
16. Tavakol M, Dennick R. Psychometric evaluation of a knowledge based examination using Rasch analysis: an illustrative guide: AMEE guide no. 72. *Med Teach*. 2013; 35:e838–48. [PubMed: 23137252]
17. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil*. 1989; 70:857–60. [PubMed: 2818162]
18. Walton D, Elliott JM. A higher-order analysis supports use of the 11-item version of the tampa scale for kinesiophobia in people with neck pain. *Phys Ther*. 2013; 93:60–8. [PubMed: 22956425]
19. Lamoureux EL, Pallant JF, Pesudovs K, Hassell JB, Keeffe JE. The Impact of Vision Impairment Questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci*. 2006; 47:4732–41. [PubMed: 17065481]
20. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951; 16:297–333.
21. Vandervelde L, Van den Bergh PY, Goemans N, Thonnard JL. ACTIVLIM: a Rasch-built measure of activity limitations in children and adults with neuromuscular disorders. *Neuromuscul Disord*. 2007; 17:459–69. [PubMed: 17433675]
22. Mayhew A, Cano S, Scott E, Eagle M, Bushby K, Muntoni F. Moving towards meaningful measurement: Rasch analysis of the North Star Ambulatory Assessment in Duchenne muscular dystrophy. *Dev Med Child Neurol*. 2011; 53:535–42. [PubMed: 21410696]
23. Berard C, Fermanian J, Payan C. Outcome measure for SMA II and III patients. *Neuromuscul Disord*. 2008; 18:593–4. author reply 4–5. [PubMed: 18595697]
24. Vanhoutte EK, Faber CG, van Nes SI, et al. Modifying the Medical Research Council grading system through Rasch analyses. *Brain*. 2012; 135:1639–49. [PubMed: 22189568]
25. Wright, BD.; Masters, GN. Rating scale analysis. Chicago: MESA Pr; 1982.

Appendix 1

The Rasch model gives estimates of questionnaire or scale item difficulty and the subjects' ability from the proportion of responses to each item on a common linear scale called “the interval scale measurement.” The measurement scale is the logit. On this scale, the items are represented by graduations or “locations” and arranged from “the easiest” to “the most difficult.” Thus, the patients are located from the least active to the most active; that is, subjects with low abilities are located on the left side of the scale (negative location), whereas subjects with high abilities are located on the right side of the scale (positive location).

For each domain of the MFM (domains D1, D2, and D3), the relevance of equally discriminant parameters was studied with a confirmatory factor analysis. A likelihood ratio test was conducted in RUMM software to determine which polytomous mathematical derivation of the Rasch model should be chosen: the Andrich's Rating Scale Model or the Masters' Partial Credit Model. Statistically significant results indicated that distances between response options varied between items to a greater extent than that expected by chance alone and, therefore, the Masters' Partial Credit Model should be used.

Fit statistics indicate how well the items fit the expected ordering required by the Rasch model. To detect misfitting, one may use either the entire response matrix (overall fit of the scale) or the individual fit (examining each item and each person). With the latter choice, the overall mean residual values for both item and person are calculated. These values are expressed as *z* scores whose mean should be close to 0 and SD close to 1. A maximum SD value of 1.5 is accepted as indicative of good fit. The summary item-trait interaction statistics reflect the fit of the observed data to the model's expectations and is represented by using the chi-square test. A significant chi-square test ($P < .05$) indicates that the data do not fit the model. In addition to the overall fit residuals, individual-item chi-square values, item residuals, and person residuals may be calculated.

To resolve the misfit, its reasons should be explored by considering the following criteria: (1) Threshold: Ordering of the response options was investigated because the MFM items have more than 2 response categories. This was done using a category probability curve for each item; this curve reflects the correct use of the MFM scoring procedure. Thresholds are said to be "disordered" when, for example, the threshold between response categories 2 and 3 is located lower on the interval scale measurement than the threshold between response categories 2 and 1. Disordered response thresholds may result from poorly worded items or ambiguous response options: this leads to an inability to reliably assign an item score. Whenever disordered thresholds were detected, item rescoring was considered according to the item's category probability curves. (2) Item bias: A potential item bias, such as the DIF, occurs when different groups within the sample (eg, men vs women, young vs elderly subjects) respond differently to an individual item despite equal levels of the underlying characteristic being measured. Every item was examined for DIF across 3 subgroups (referred to in RUMM as "person factors"): age (<18y vs 18y), sex (males vs females), and diagnosis (CM vs CMD). To assess DIF in RUMM, an analysis of variance of the standardized response residuals for each item was used at different levels of each trait. A Bonferroni-adjusted alpha level was used to determine the statistical significance. In addition, the importance of DIF was assessed graphically. Items that exhibited DIF (statistically and graphically) were considered for removal and were actually removed when their removal improved the overall model fit. (3) Location dependence: This occurs when the response to a given item is influenced by the response to a previous item (eg, questions about walking followed by questions about running abilities). To identify location dependence, the residual correlation matrix generated in RUMM was examined and pairs of items with correlations exceeding 0.3 were considered dependent and sources of misfit. To test the effect of location dependence on the model fit, a subtest procedure is performed in which conceptually similar items are grouped to form a super item and reentered into the Rasch analysis. When this procedure corrects the misfit, the scale is said to conform

adequately to the Rasch measurement model. This was applied to MFM domains D2 and D3.

Once the degree of fit of the data to the Rasch model is determined by the appropriate range of statistics, it is necessary to confirm that the scale is appropriately targeted to the population being assessed. This is done by comparing the mean location score for person to 0 (0 being the mean location score for the items). With a well-targeted measure (ie, not too easy and not too hard), the mean location for persons, as indicated by the person-item threshold distributions, would be around 0. A negative mean value indicates that the sample, as a whole, is located at a lower level than the average (floor effect), whereas a positive value would suggest the opposite (ceiling effect).

To confirm the unidimensionality of the scale, we used a principal-component analysis of the residuals—available in RUMM 2030—to identify the 2 most distant subsets of items; that is, identify items with positive and negative signs. These 2 sets of items were then used to obtain separate person estimates. Using an independent *t* test for the difference in these estimates for each person, the percentage of significant tests outside the mean ± 1.96 SD range should not exceed 5%.

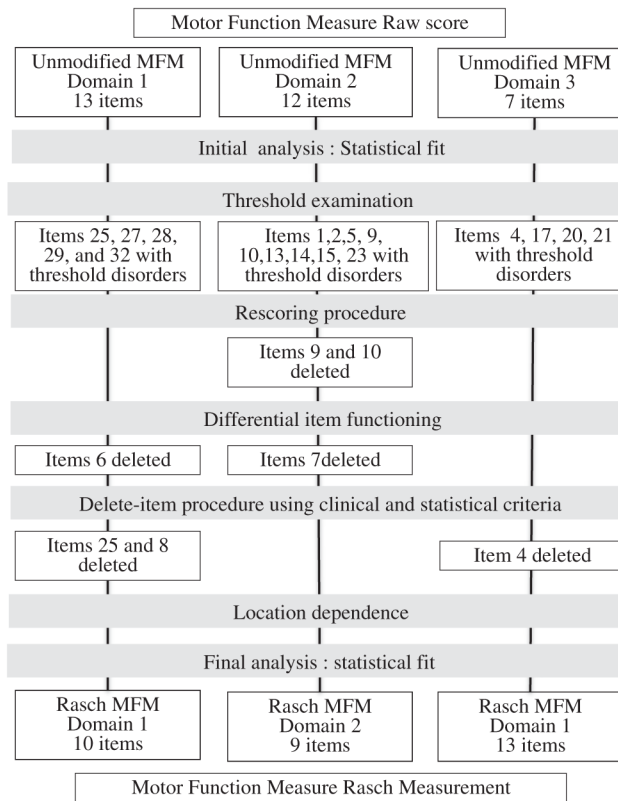


Fig 1.
Flow chart of the Rasch analysis.

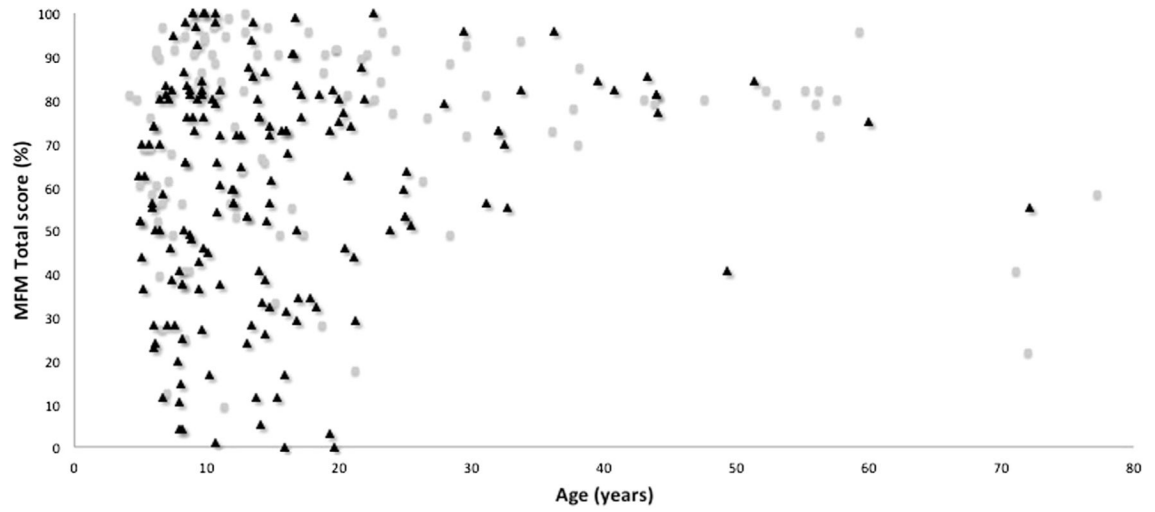


Fig 2. Thirty-two-item MFM total score according to age in 98 patients with CMs (hollow dots) and 191 patients with CMD (filled triangles).

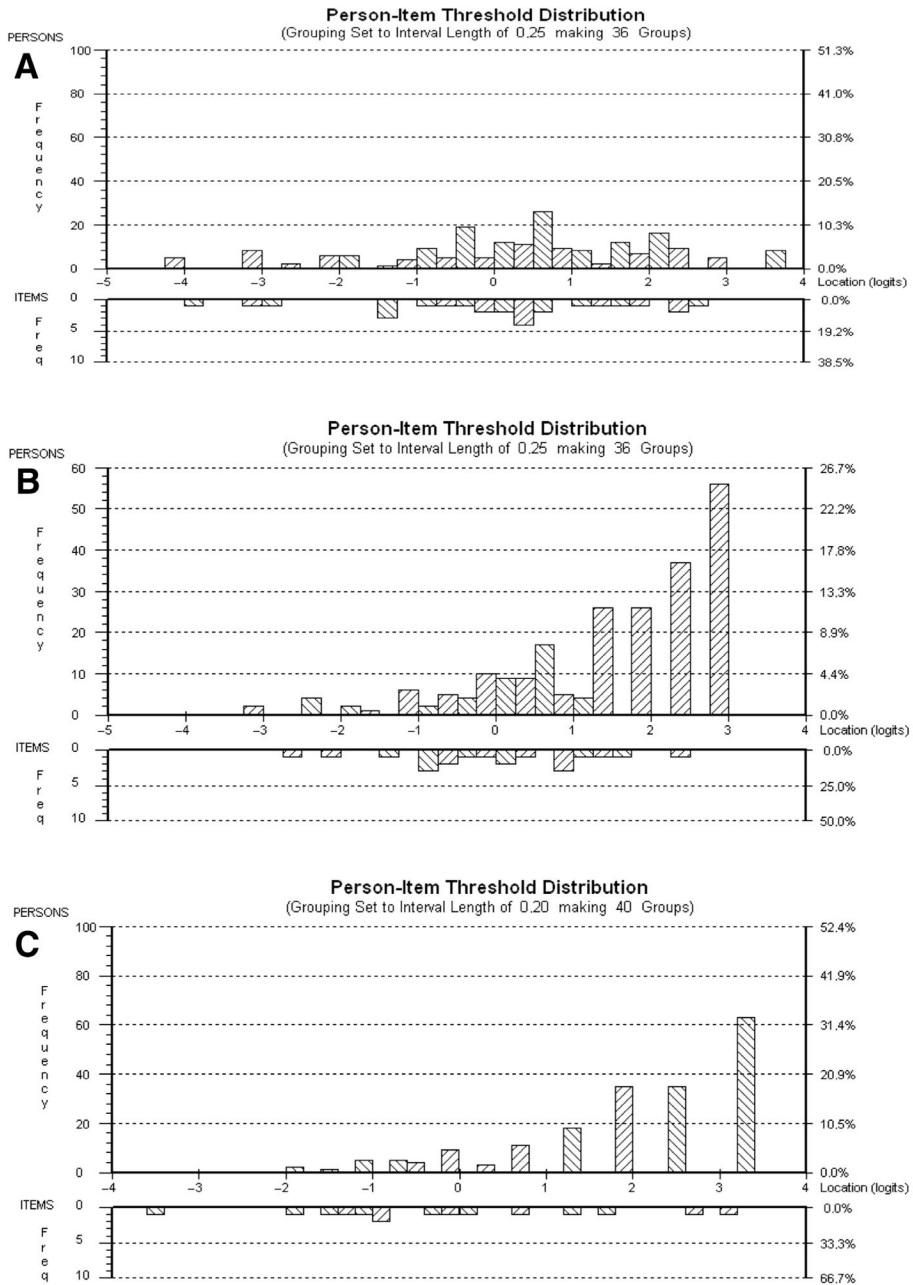


Fig 3. Person-item threshold histograms for Rasch-scaled MFM domains D1 (A), D2 (B), and D3 (C). Participants with the highest level of motor function and the most difficult items are on the right side, whereas those with the lowest level of motor function and the least difficult items are on the left side.

Main patients' characteristics and MFEM scores

Table 1

Diagnosis	N	Age (y)	M/F	Ambulant (%)	13-Item D1	MFEM Score				
						12-Item D2	7-Item D3	32-Item total		
CMs	98	21.1±17.2	43/56	75.9	57.9±30.7*	84.8±19.5	90.5±12.3	75.2±20.1		
Nemaline myopathy	15	25.6±23.4	4/11	60	50.4±36.3	82.6±20.6	87.9±11.4	70.7±22.5		
Centronuclear myopathy	10	28.4±15.0	3/7	90	55.9±23.1	82.8±24.7	87.6±20.7	72.9±22		
Central core disease	20	22.8±17.6	9/11	78.9	59.2±30.4	89.1±19	94.3±9.6	78.1±19.2		
Multiminicore disease	13	16.1±10.3	8/5	84.6	56±32.6	81.2±17.6	91.6±9.2	73.2±19.1		
Other CM	40	17.9±16.1	18/22	84	59.3±29.1	84.5±17.9	90±12.1	75.5±18.9		
CMD	191	14.8±11.0	86/83	56.8	37.2±32.3	73.1±27.1	81.6±20.4	60.5±24.8		
Laminin α 2-deficient CMD	43	12.2±7.7	16/20	48	21.3±29.9	62.1±32.9	71.1±26.2	49.7±28.3		
Collagen VI-related CMD	100	16.1±12.0	46/48	58.3	41.4±31.5	78.6±23.1	87.1±15.2	64.6±21.7		
Abnormal glycosylation of dystroglycan	8	12.8±6.6	4/4	62.5	54.2±35.1	74±34.6	72.6±30.4	65.6±32.6		
Other CMD	40	17.4±11.3	22/16	57.9	39.6±31.7	72.8±26.2	84.1±17.6	61.8±23.8		
Whole sample	289	17.4±13.9	129/139	59.75	44.1±33.3	77.4±25.4	85.1±18.6	65.6±24.3		

NOTE. Values are mean ± SD unless indicated otherwise.

Abbreviations: F, female; M, male.

* Score ± SD expressed as the percent of maximum possible score in each MFEM domain and total score.

Table 2
Values of the criteria of the remaining 28 items after examination of the thresholds and the DIF

MFIM Domains	Item	Location	Fit Residuals	Probability Score	Ceiling Effect*	Deviation [†]	Removal Agreement [‡] (%)	Item Difficulty [§]
D1 domain	25	Maintain standing position	3.243//	<0.001//	0.49	Yes	50	3.1 (1-5)
	12	Sit down from standing	-0.438	0.00	0.41	No	0	3.0 (3-5)
	26	Raise the foot	-0.88	0.07	0.41	No	0	4.0 (2-5)
	29	Takes 10 steps on a line	-0.35	0.09	0.48	No	33.3	3.8 (2-5)
	8	From supine, sit up	5.462//	<0.001//	0.15	Yes	0	4.0 (4-5)
	24	Stand up from the mat	0.662	0.00	0.25	No	0	4.0 (3-5)
	27	Touch the floor	-0.83	0.05	0.36	No	0	4.6 (3-5)
	11	Stand up from the chair	0.579	0.68	0.13	No	0	5.0 (4-5)
	28	10 steps on both heels	-1.017	0.02	0.19	Yes	50	5.0 (4-5)
	30	Run 10 meters	-3.376//	0.03	0.1	No	0	4.8 (4-5)
	32	Squat from standing	-1.678	0.04	0.16	No	0	4.8 (4-5)
	31	Hop 10 times in place	-1.641	0.11	0.11	No	0	5.0 (5)
	23	Place the forearms on the table	-2.125	0.13	0.83	No	0	2.0 (0-3)
D2 domain	16	Extend the elbow	-0.342	0.37	0.64	No	50	2.0 (2)
	1	Hold the head for 5 seconds	-1.788	0.07	0.71	No	83.3	0.5 (0-1)
	15	Place hands on the head	-1.762	0.00	0.54	No	33.3	3.0 (2-4)
	5	Move hand to opposite shoulder	-2.289	0.01	0.7	No	0	1.2 (0-3)
	3	Flex hip and knee >90°	1.628	0.45	0.59	Yes	50	2.6 (1-3)
	13	Maintain the seated position	0.749	0.72	0.58	No	0	2.1 (1-3)
	14	Head in flexion, raise the head	-0.188	0.10	0.62	Yes	66.6	2.0 (1-3)
	2	From supine, raise the head	1.217	0.06	0.25	No	33.3	4.0 (3-5)
	22	Place the finger on 8 drawings	-1.957	0.07	0.82	No	83.3	0 (0-2)
D3 domain	17	Pick up and hold 10 coins	-2.435	0.03	0.61	No	33.3	0 (0-2)
	18	Go round the edge of the digital video disc	-2.486	0.02	0.74	No	33.3	2 (1-4)
	4	Dorsiflex the foot	3.489//	<0.001*	0.54	Yes	50	1 (0-2)
	21	Pick the ball and turn the hand	-0.91	0.19	0.68	No	33.3	2 (0-3)
	19	Draw a series of loops	1.263	0.09	0.41	No	33.3	2 (1-2)

MFIM Domains	Item	Location	Fit Residuals	Probability Score	Ceiling Effect*	Deviation [†]	Removal Agreement [‡] (%)	Item Difficulty [§]
20	Tear the sheet of paper folded	1.524	-1.641	0.07	0.35	No	33.3	2 (2-3)

* Percentage of patients with 0 score.

[†] Graphical examination of item characteristic curves for significant deviations from the expected model.

[‡] Percentage of agreement between 6 physiotherapists to change or remove the item.

[§] Median (range) of item difficulty from 0 (very easy) to 5 (very difficult) as assessed by physiotherapists.

// Fit residuals >2.5 or < -2.5.

¶ Probability below the Bonferroni adjustment for D1 (.00077 for 13 items), D2 (.00083 for 12 items), and D3 (.00143 for 7 items).

Summary of Rasch analyses statistics relative to the MFEM domain scores in patients with congenital disorders of muscle

Table 3

MFEM Domains	Item Location	Item Fit Residuals	Person Location	Person Fit Residuals	Item-Trait χ^2 Interaction			Unidimensionality Independent <i>t</i> Test (%)
					<i>df</i>	<i>P</i>	PSI	
D1 domain								
Initial analysis (13 items)	0.000±1.003	-0.459±1.806	-0.248±2.016	-0.379±0.905	52	<.0001	.92	.085
Final analysis (10 items)	0.000±1.115	0.213±1.369	0.371±1.722	-0.250±0.919	40	.09	.89	.059
D2 domain								
Initial analysis (12 items)	0.000±0.731	-0.538±1.560	1.115±1.171	-0.292±0.848	48	<.0001	.82	.027
Final analysis (9 items)	0.000±0.538	-0.633±1.244	1.632±1.496	-0.376±0.817	21	.08	.77	.028
D3 domain								
Initial analysis (7 items)	0.000±1.186	-0.798±2.202	1.547±1.083	-0.320±0.739	21	<.0001	.59	.014
Final analysis (6 items)	0.000±1.186	-0.155±0.708	2.358±1.621	-0.326±0.707	10	.29	.67	.017

NOTE. Values are mean ± SD.

Abbreviation: PSI, Person separation index.