RESEARCH ARTICLE

Multicenter Validation of the Ocular Myasthenia Gravis Rating Scale Questionnaire

Sui Hsien Wong, Rebhi Abuzaitoun, Wayne T. Cornblath, Eric R. Eggenberger, Sangeeta Khanna, Tatiana Deveney, Joshua R. Ehrlich, Chris A. Andrews, Carolina Barnett-Tapia, and Lindsey B. De Lott

Correspondence
Dr. Wong
suiwong@doctors.org.uk

Neurology® 2025;104:e210150. doi:10.1212/WNL.0000000000210150

Abstract

Background and Objectives

Ocular myasthenia gravis (OMG) causes disabling ocular symptoms of ptosis and diplopia, but a validated disease-specific patient-reported outcome measure (PROM) has not been reported. We sought to validate a novel PROM for OMG, OMG Rating Scale Questionnaire (OMGRate-q), as a measure of visual functioning to support patient-centered care.

Methods

This was a prospective study of patients aged 18 years and older with OMG receiving care at 3 medical centers (January 2022–October 2023). The 10-item OMGRate-q was administered, and response data were analyzed using exploratory factor analysis followed by Andrich rating scale model fitting. Poorly fitting items were eliminated, and the model was refit to produce the final items, item locations, and thresholds. Latent scores (theta) were estimated, test-retest reliability was established with repeat measures, and correlation with other myasthenia gravis PROMs was measured.

Results

Of the 134 patients included in the study, 45 (33.6%) were women, 99 (73.9%) were White, and the median age (interquartile range [IQR]) was 64.6 years (52.6–73.9 years). A ptosis-related item showed significant item-trait deviation ($p < 0.001$) and was kept as a separate factor from the remaining diplopia-related items. After excluding this item, there were no misfitting items. Theta estimation for the diplopia scale ranged from −3.47 to 5.51 with median = −0.53 (IQR −2.33, 0.72). Test-retest reliability of the OMGRate-q diplopia subscale was high (intraclass correlation coefficient = 0.95 [95% CI 0.90–0.98]) and of the ptosis item was good (weighted $\kappa = 0.56$). No significant differences were observed in OMGRate-q diplopia subscale scores or the ptosis item between the 3 sites (diplopia $p = 0.44$; ptosis $p = 0.32$). OMGRate-q scores were moderately to highly correlated with the Myasthenia Gravis Quality of Life 15 questionnaire (n = 122; diplopia: $r = 0.68$, $p < 0.001$; ptosis: $r = 0.48$, $p < 0.001$) and Myasthenia Gravis Impairment Index (n = 130; diplopia: $r = 0.76$, $p < 0.001$; ptosis: $r = 0.77$, $p < 0.001$). OMGRate-q length was acceptable to most participants (125/130 [96.2%]), and the questionnaire was completed in 80.7 (±45.2) seconds.

Discussion

The OMGRate-q is a valid and reliable disease-specific PROM for OMG that may be used to facilitate patient-centered research and care. However, the OMGRate-q emphasizes the impact of diplopia on visual functioning with a single separate item measuring ptosis. Future studies are needed to determine OMGRate-q responsiveness to disease-state changes and how to add measures of ptosis to this scale or whether a separate measure is needed.

From the Department of Neuro-Ophthalmology (S.H.W.), Moorfields Eye Hospital NHS Foundation Trust; Departments of Ophthalmology and Neurology (S.H.W.), Guy's and St Thomas' NHS Foundation Trust; Department of Ophthalmology (S.H.W.), Faculty of Life Sciences & Medicine, King's College London; Department of Clinical and Movement Neuroscience (S.H.W.), UCL Queen Square Institute of Neurology, University College London, United Kingdom; Department of Ophthalmology and Visual Sciences (R.A., W.T.C., T.D., J.R.E., C.A.A., L.B.D.L.), Kellogg Eye Center, University of Michigan, Ann Arbor; Departments of Ophthalmology, Neurology, and Neurosurgery (E.R.E., S.K.), Mayo Clinic, FL; Survey Research Center (J.R.E.), Institute for Social Research, University of Michigan, Ann Arbor; and Division of Neurology (C.B.-T.), Department of Medicine, University Health Network and University of Toronto, Ontario, Canada.

Go to Neurology.org/N for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Glossary

AUROC = area under the ROC curve; **DIF** = differential item functioning; **EFA** = exploratory factor analysis; **ICC** = intraclass correlation coefficient; **IQR** = interquartile range; **IRB** = institutional review board; **MG** = myasthenia gravis; **MGII** = Myasthenia Gravis Impairment Index; **MG-QOL-15** = Myasthenia Gravis Quality of Life 15 questionnaire; **OMG** = ocular myasthenia gravis; **OMGRate-q** = OMG Rating Scale Questionnaire; **PASS** = patient acceptable symptom state; **PROM** = patient-reported outcome measure; **ROC** = receiver operating characteristic.

Introduction

Ocular myasthenia gravis (OMG) causes potentially disabling ocular symptoms of ptosis and diplopia. Capturing the patient experience of living with OMG using patient-reported outcome measures (PROMs) is a recommended part of delivering patient-centered care. Our previous work has demonstrated that existing myasthenia gravis (MG)–specific PROMs and visual quality-of-life measures are insufficient for measuring the impact of ocular symptoms on the daily lives of patients with OMG.[1] The lack of a disease-specific OMG PROM is a significant barrier to designing clinical trials addressing the OMG and delivering patient-centered OMG care. Therefore, we developed and preliminarily validated a 10-item rating scale, the Ocular Myasthenia Gravis Rating Scale Questionnaire (OMGRate-q).[2]

Our preliminary work established that the OMGRate-q was suitable for further validation. In a single-center study, we observed excellent correlation between the OMGRate-q and Myasthenia Gravis Impairment Index (MGII) ocular subscale (N = 104 patients, $r = 0.85$, 95% CI 0.78–0.91, $p < 0.001$), which has been shown to capture important constructs of symptom fluctuation and severity. We also observed good correlation of the OMGRate-q with the Myasthenia Gravis Quality of Life 15 questionnaire (MG-QOL-15; $r = 0.71$, 95% CI 0.64–0.79, $p < 0.001$), suggesting that the OMGRate-q reflects symptoms that affect the quality of life in patients with OMG. A focus group of 15 patients found that the OMGRate-q was usable and preferred over the MGII and the MG-QOL-15.[2] In addition, patients perceived that the MGII and MG-QOL-15 did not sufficiently capture the patient experience of living with OMG. In this multicenter study, we aimed to further validate and establish the reliability of the OMGRate-q as a disease-specific PROM for patients with OMG.

Methods

Study Population

We prospectively collected data from patients with OMG receiving care at Moorfields Eye Hospital, University of Michigan's Kellogg Eye Center, and Mayo Clinic Florida from January 17, 2022, to October 19, 2023. Eligible participants were aged 18 years or older with a diagnosis of OMG confirmed by a fellowship-trained neuro-ophthalmologist (S.H.W, L.B.D.L., W.C., E.E, S.K., T.D.). The diagnosis was made by the presence of variability of ptosis and/or diplopia and a supportive paraclinical test. Supportive testing included one or more of the following: positive antibodies (acetylcholine receptor, muscle-specific kinase, low-density lipoprotein receptor 4), changes consistent with a neuromuscular junction disorder on single-fiber or repetitive stimulation electromyography, or positive ice test.[3] Patients were also eligible if they had diplopia and/or ptosis with negative paraclinical testing but objective clinical improvement in ocular misalignment or ptosis in response to pyridostigmine. Patients who did not speak English or had a diagnosis of cognitive impairment were excluded. In addition, patients with muscle weakness in non-ocular muscle groups were also excluded.

Data Collection

Questionnaires were administered either during an in-person encounter or by an examiner in a video encounter. During in-person encounters, paper questionnaires were completed by patients. In video encounters, questionnaires were completed by the clinician. Repeat measures were also collected by mail. Clinical records including demographic information and medical history were collected from the electronic health records for analysis. These data included age, sex, race, ethnicity, serologic testing results, neurophysiologic testing results, and use of potentially therapeutic medications (pyridostigmine, corticosteroids, and nonsteroidal immunosuppression). Sex, race, and ethnicity were self-reported.

Standard Protocol Approvals, Registrations, and Patient Consents

This study received institutional review board (IRB) approval at the 3 respective sites (Moorfields Eye Hospital, University of Michigan, Mayo Clinic Florida), where the ethics boards determined that participant consent was not required (Moorfields Eye Hospital (IRB no. 1100), University of Michigan (IRB no. HUM00161470)) or written informed consent was obtained from the participants in the study (Mayo Clinic Florida, IRB no. 19-011231), respectively.

Survey Items

We previously developed the OMGRate-q (in eAppendix 1) through an iterative process based on expert opinion (S.H.W, E.E, W.C.) and patient interviews.[2] The OMGRate-q adapted 10 relevant questions from validated questionnaires including the Neuro-ophthalmic Supplement of the National Eye Institute Visual Function Questionnaire,[4] the Adult Strabismus Quality of Life Questionnaire,[5] and the Diplopia Questionnaire.[6] All questions have 5 response options from

low to high: never, rarely, sometimes, often, and always. The recall period was 2 weeks. We selected this time interval based on previous work performed for the MGII,[7] which demonstrated that this is a useful timescale to capture the fluctuations of symptoms in generalized MG. The questionnaire was written at a Flesch-Kincaid sixth-grade US reading level.

Since the original publication of the OMGRate-q, 2 minor changes were made to the scale. First, question 3 ("I have difficulty judging distances [i.e., problems with depth perception]") was split into 2 questions ("I have difficulty judging distances" and "I have problems with depth perception") because patients may conceptualize changes in depth perception and judging distances differently. Second, we removed a redundant question about the presence of diplopia ("in the past two weeks, have you had double vision looking any direction or at any time of the day?") because subsequent questions query the presence/absence of diplopia in different positions of gaze, with response options of "never, rarely, sometimes, often, always."

Data Analysis

Descriptive statistics were used to summarize participant demographics and characteristics.

Scale Dimensionality and Factor Analysis

Exploratory factor analysis (EFA) was performed to assess the dimensionality of the OMGRate-q and whether all items were loading on a unique factor or multiple factors (e.g., diplopia and ptosis). To assess the number or factors, we used principal axis factoring with parallel analysis. This simulation-based method tests the probability of the factor being generated because of chance, aims to minimize overidentification of factors, and is superior to relying solely on eigenvalue scores generated by factor analysis alone.[8] Oblique (Promax) was the method of rotation. Extraction was performed using the polychoric/tetrachoric correlation matrix, and factor loadings above 0.40[9] were displayed. The number of factors was determined based on visualization of the scree plot.

Rasch Model

We planned to fit a Rasch model to the measure if the EFA was determined to be unidimensional; in the case that more than 1 factor was found, we planned to use a Rasch model for all factors with more than 3 items. We used the Andrich rating scale approach based on our sample size and its ability to consider polytomous response categories. Fitting the model was performed using conditional maximum likelihood estimation. This model assigns 1 location for each item and 4 spacings between the 5 response levels (common to all items). Participants who had missing responses on any of the items were excluded from model fitting because the statistical software package used would omit all cases with missing item responses. Participants with missing responses were later included in score estimation.

Item fit measures how each of the items deviates from measuring the latent score. Item-trait interaction was measured using the χ^2 test, where a nonsignificant value indicates overall good fit. Total item-trait interaction was estimated for the model.[10] For individual items, we estimated INFIT and OUTFIT of the mean square. A value between 0.6 and 1.4 was allowed for INFIT and OUTFIT mean square because these cutoffs are considered reasonable for the rating scale.[11] We also assessed the order of thresholds of each items, through plots. Items that showed item-trait misfit were removed one at a time, and the model was refit with the remaining items until we detected no significant item-trait deviation.

Latent score estimation (theta) was performed by the anchored maximum likelihood estimation method.[12] 1 theta is equivalent to 1 SD from the mean latent score of the entire sample with a higher theta indicating worse disease. Participants with missing item responses were included. We provided a percentile metric for ease of interpretation of latent scores. A person-item map was constructed to show the thresholds and location for each item.

Differential Item Functioning and Local Dependency

Differential item functioning (DIF) occurs when an external variable, such as age, affects the latent score (theta) measured with the instrument. Presence of DIF was evaluated for age (above vs below median), country, and sex. We assessed DIF through subtracting item locations, through comparing means of item residuals using the Welch t test where a significant p value indicates significant DIF, and visually through item plots to assess for uniform and nonuniform DIF.

Local dependency measures the correlations among items after extracting the effect of the latent trait being measured. For each item, residuals from subtracting actual and expected responses were obtained and then correlated across items. A Pearson correlation coefficient of 0.7 was considered as the maximum tolerated value.[13]

Reliability, Trait-Variable Associations, and Construct Validity

Test-retest variability is a measure of precision between the initial and repeated measures. A subset of 25 patients completed repeated measures with a separation time of 14–28 days.[14] We estimated Bland-Altman limits of agreement and intraclass correlation coefficients (ICCs; 2-way random effect, absolute agreement, multiple raters) for the diplopia domain score. For individual item reliability, we used weighted κ, which was also used to evaluate test-retest reliability for the ptosis subscale. Discriminative ability was measured by person separation reliability, and a value of $p \geq 0.7$ was acceptable.[15]

Trait-variable associations of both domains were assessed by Spearman ρ for association with the following: age, sex, and seropositivity (dichotomous). Scores between the 3 different sites were compared using Welch analysis of variance.

Associations between the OMGRate-q and the MG-QOL-15 and MGII (6-item eye-related subscale only) were evaluated using Pearson correlation. We assessed the correlation between the OMGRate-q score and the ocular-focused item in MG-QOL-15 (item 2) through Spearman ρ. Based on our previous study, we hypothesized that the OMGRate-q would be strongly correlated with the MG-QOL-15 and MGII. Missing values on the MGII were imputed using the mean imputation as per development instructions. Participants with missing responses on all MGII or MG-QOL-15 items were excluded.

Interpretability

Patient acceptable symptom state (PASS) identifies the highest level of disease symptoms with which patients are comfortable. Symptom intensities below the PASS score are considered acceptable.[16] We used a previously validated question in MG,[17] which we slightly modified for OMG: "Considering all the ways your Ocular Myasthenia symptoms have affected you during the last month, would you deem it acceptable if you remained at a similar level of Ocular Myasthenia symptom severity for the next few months?" We used logistic regression receiver operating characteristic (ROC) curve analysis to identify the OMGRate-q subscale scores where the PASS question is equally likely to be answered "Yes" or "No."

Scale Acceptability and Feasibility

Patient acceptability of the OMGRate-q was assessed using 3 questions. The first was a 3-level Likert scale asking whether the length of OMGRate-q was "Too short," "Too long," or "Just Right." The second question was an open-ended inquiry asking participants about what they liked about the OMGRate-q. The third question asked participants for suggestions related to improving the questionnaire. Feasibility of OMGRate-q use in clinical environments was measured using the time to complete the questionnaire in a subset of patients.

Statistical Software

JASP statistics software (version 0.18.2; University of Amsterdam, Amsterdam, the Netherlands) was used for EFA. For model fitting and statistics, we used the configuration of eRm package[18] in R 4.4.0 (R Foundation for Statistical Computing, Vienna, Austria)[19] and SPSS 28 (IBM, Armonk, NY).

Data Availability

Anonymized data sets are available on reasonable request.

Results

Of the 134 patients included in the study, 89 (66.4%) were men. The median age (interquartile range [IQR]) was 64.6 years (52.6–73.9 years), and most (n = 99, 73.9%) were White. Most questionnaires were administered during an in-person visit encounter (n = 129, 96.3%). Most patients (n = 97, 72.4%) were seropositive. The median (IQR) disease duration was 4.1 (1.3, 6.4) years. Table 1 provides participant characteristics overall and by serologic status. No notable differences were observed between the seropositive and seronegative patients. eTable 1 provides characteristics among seropositive patients by antibody type.

Scale Dimensionality, Factor Analysis, Model Fit, and DIF

All 10 items were loaded on a single factor. The least loading value was 0.56, and the highest was 0.91 (eTable 2). The scree plot (eFigure 1) supported the unidimensionality of the measure. Of the 134 participants, 5 had at least 1 missing response (eTable 3) and were excluded from Rasch model fitting (129 participants for model fitting). Initially, all 10 items were fit into the model. All items had good item fit except item 1, "My eyelid (or eyelids) droop due to my Ocular Myasthenia Gravis," which showed significant item-trait deviation ($p < 0.001$). This item was excluded, and further analyses refer to the remaining diplopia-related items. The ptosis item had a median (IQR) score of 1 (0–3).

After refitting the model, there were no significant item-trait deviations (Table 2). The final model for the diplopia scale had good total-item trait interaction ($\chi^2 = 856.5$, $df = 921$, $p = 0.67$). All items had ordered thresholds (eFigure 2). Final item location, fit statistics, and thresholds are presented in Table 3. We also provide a percentile reference for ranking the latent trait estimation in Table 4. None of the items had a Pearson correlation coefficient higher than 0.5 indicating absence of significant local dependency when adjusting for the latent trait being measured (Table 5).

Theta was calculated for all patients. No more than 6 iterations were needed for theta estimation. The smallest observed value was −3.47 while the largest was 5.51. The median (IQR) was −0.53 (−2.33 to 0.72). The person-item map is presented in Figure 1. DIF based on age and country was not observed through subtracting item locations (eTable 4). Sex could not be assessed using the first DIF method because only 45 participants were women, which was less than the minimum needed for fitting the model. There was no difference in mean item residuals based on age, sex, or country, indicating absence of significant DIF (Table 6 and eFigure 3).

Reliability, Trait-Variable Associations, Construct Validity, and Interpretability

The person separation reliability was 0.85 indicating good discriminative ability. Bland-Altman lower and upper limits of agreement for test-retest variability on the diplopia subscale were −1.44 and 1.63 (eFigure 4). The mean change in the diplopia subscale score (95% CI) between the initial and repeated measure was 0.09 (−0.23 to 0.42), with a high ICC of 0.95 (95% CI 0.90–0.98). Individual item reliability statistics are given in eTable 5. As noted in eTable 4, the ptosis subscale (item 1) had good test-retest reliability ($\kappa = 0.56$).

The associations of the OMGRate-q diplopia subscale score with age ($\rho = 0.04$, $p = 0.65$), sex ($\rho = 0.17$, $p = 0.05$), and

Table 1 Participant Demographics and Characteristics (N = 134)

Characteristic	Overall	Seropositive	Seronegative[a]
Sample, n	134	97	37
Age, y, median (IQR)	64.6 (52.6–73.9)	66.1 (55.4–74.3)	56.5 (47.5–71.4)
Sex, n (%)			
Male	89 (66.4)	64 (66.0)	25 (67.6)
Female	45 (33.6)	33 (34.0)	12 (32.4)
Race, n (%)			
White	99 (73.9)	70 (72.2)	28 (75.7)
African American, Afro-Caribbean, Black African, or Black British	19 (14.2)	16 (16.5)	3 (8.1)
Asian	15 (11.2)	10 (10.3)	5 (13.5)
Unknown	1 (0.8)	1 (1.0)	1 (2.7)
Ethnicity, n (%)			
Non-Hispanic	129 (96.3)	93 (95.9)	36 (97.3)
Hispanic	3 (2.2)	3 (3.1)	0 (0.0)
Unknown	2 (1.5)	1 (1.0)	1 (2.7)
Site, n (%)			
Moorfields Eye Hospital	79 (59.0)	56 (57.7)	23 (62.2)
University of Michigan	39 (29.1)	25 (25.8)	14 (37.8)
Mayo Clinic	16 (11.9)	16 (16.5)	0 (0.0)
Disease duration, y, median (IQR)	4.1 (1.3–6.4)	4.5 (2.7–6.5)	3.9 (1.0–6.4)
Definitive diagnostic test, n (%)			
Acetylcholine receptor	87 (64.9)	87 (84.4)	—
Muscle-specific kinase	8 (6.0)	8 (8.3)	—
Low-density lipoprotein receptor–related protein 4	2 (1.5)	2 (2.1)	—
Positive single-fiber electromyography[b]	24/28 (85.7)	—	24/28 (85.7)
Positive response to pyridostigmine[b]	21/36 (58.3)	—	21/36 (58.3)
Therapeutic medication, n (%)[c]			
Pyridostigmine	76 (58.0)	55 (56.7)	21 (56.8)
Corticosteroids	35 (27.0)	28 (28.9)	7 (18.9)
Steroid-sparing immunosuppression	20 (15.2)	12 (12.4)	8 (21.6)
No medication	22 (16.4)	16 (16.5)	6 (16.2)
Survey administration (initial), n (%)[d]			
In-person	129 (96.3)	94 (96.9)	35 (94.6)
Video encounter	4 (3.0)	3 (3.1)	1 (2.7)
Mail	1 (0.7)	0 (0.0)	1 (2.7)

Abbreviation: IQR = interquartile range.
[a] Diagnosis of seronegative ocular myasthenia gravis was based on significant clinical response to pyridostigmine or a finding consistent with a neuromuscular junction disorder on single-fiber electromyography.
[b] Missing values: positive single-fiber electromyography (inclusive of repetitive stimulation) = 9; positive response to pyridostigmine = 1.
[c] Missing values: pyridostigmine = 3; corticosteroids = 5; steroid-sparing immunosuppression = 2.
[d] Repeat measures: n = 22 of 25 (88%) were administered by mail; 3 (12%) were in-person.

Table 2 Item Fit Statistics for the Diplopia Subscale

Item number	Brief item description	χ^2	*df*	*p* Value	OUTFIT MSQ	INFIT MSQ	OUTFIT *t*	INFIT *t*
2	Close 1 eye to improve vision	101.5	102	0.49	1.0	1.0	0.0	0.1
3	Judging distance difficulty	125.3	103	0.07	1.2	1.0	1.1	−0.3
4	Depth perception problems	116.2	102	0.16	1.1	0.9	0.7	−0.4
5	Reading[a]	77.4	99	0.95	0.8	0.8	−1.2	−1.2
6	Looking in distance[a]	66.6	103	1.00	0.6	0.7	−2.1	−1.9
7	Looking up[a]	75.3	103	0.98	0.7	0.7	−1.6	−2.1
8	Looking down[a]	73.7	103	0.99	0.7	0.8	−1.6	−1.3
9	Looking right[a]	114.3	103	0.21	1.1	1.2	0.6	1.4
10	Looking left[a]	106.2	103	0.39	1.0	1.0	0.2	0.3

Abbreviations: *df* = degree of freedom; MSQ = mean square.
p Values are obtained from the χ^2 value and degree of freedom.
[a] Item asks about double vision in specific gazes and with specific tasks.

seropositivity ($\rho = 0.00$, $p = 1.00$) were not statistically significant. No significant differences were observed in diplopia scores between the 3 sites ($p = 0.44$). OMGRate-q scores and MG-QOL-15 (n = 122, $r = 0.68$, $p < 0.001$) and MGII (n = 130, $r = 0.76$, $p < 0.001$) items were moderately to highly correlated with the diplopia domain. The ptosis subscale showed similar results, no association with age ($\rho = -0.03$, $p = 0.71$), sex ($\rho = 0.11$, $p = 0.19$), and seropositivity ($\rho = -0.08$, $p = 0.34$). No significant differences were observed between the 3 sites ($p = 0.32$), and a significant correlation with MG-QOL-15 (n = 122, $r = 0.48$, $p < 0.001$) and MGII (n = 130, $r = 0.77$, $p < 0.001$) was found. A significant correlation was also found between the ocular-focused item in MG-QOL-15 (item 2) and OMGRate-q score ($\rho = 0.78$, $p < 0.001$).

Of 128 patients who answered the dichotomized global PASS question, 32 (25.0%) indicated that the symptom level was unacceptable. The diplopia subscale value where there is equal probability of the disease symptoms being acceptable or not acceptable was at 0.04 (area under the ROC curve [AUROC] = 0.84; sensitivity = 0.78; specificity = 0.73). Therefore, an OMGRate-q diplopia subscale value higher than 0.04 (61st percentile) would indicate that the OMG diplopia symptoms are likely *not* acceptable (eFigure 5). As for the ptosis domain, equal probability of the disease symptoms being acceptable or not acceptable was at the response category "sometimes" (AUROC = 0.80; sensitivity = 0.84; specificity = 0.75), indicating that a response of the "often" or "always" categories indicated unacceptable symptoms.

Scale Acceptability and Feasibility

Most participants found OMGRate-q length "just right" (125/130 [96.2%]). Of the 79 participants who provided qualitative feedback about the OMGRate-q, 78 (98.7%)

Table 3 Item Locations and Threshold Parameters for the Diplopia Subscale

Item number	Brief item description	Location	Threshold 1	Threshold 2	Threshold 3	Threshold 4
2	Close 1 eye to improve vision	0.311	−0.644	−0.629	1.097	1.418
3	Judging distance difficulty	0.953	−0.002	0.013	1.740	2.060
4	Depth perception problems	1.243	0.289	0.303	2.030	2.351
5	Reading[a]	1.120	0.166	0.181	1.907	2.228
6	Looking in distance[a]	1.108	0.153	0.168	1.894	2.215
7	Looking up[a]	1.108	0.153	0.168	1.894	2.215
8	Looking down[a]	1.180	0.225	0.240	1.966	2.287
9	Looking right[a]	0.711	−0.243	−0.229	1.498	1.818
10	Looking left[a]	0.857	−0.097	−0.082	1.644	1.965

[a] Item asks about double vision in specific gazes and with specific tasks.

Table 4 Percentile Metric for Latent Scores (Theta) on the Diplopia Subscale

Percentile	Theta (OMGRate-q diplopia subscale score)
20	−3.48
25	−2.33
30	−1.67
35	−1.30
40	−1.02
45	−0.80
50	−0.53
55	−0.18
60	−0.03
65	0.10
70	0.46
75	0.72
80	1.04
85	1.30
90	1.59
95	2.24
100	5.51

responded positively. Participants indicated that it "addressed symptoms," "covered all concerns," was "easy to do," had "clear boxes," and "enables the doctors to know how best to treat the problem." Most participants (56/72, 77.8%) had no suggestions for improving the instrument. Suggestions for improvement included the OMGRate-q asking about the following: whether closing 1 eye helps with the double vision, presence of blurry vision, the last ocular episode (instead of the last 2 weeks), using 4 weeks instead of 2 for the symptom duration, allowing for comments alongside the 5 answer options, implementing an online version and an app, adding a section for times of occurrence of symptoms, and inquiring whether the condition affected the feelings of participants. The mean (SD) time to complete the OMGRate-q was 80.7 (±45.2) seconds.

Discussion

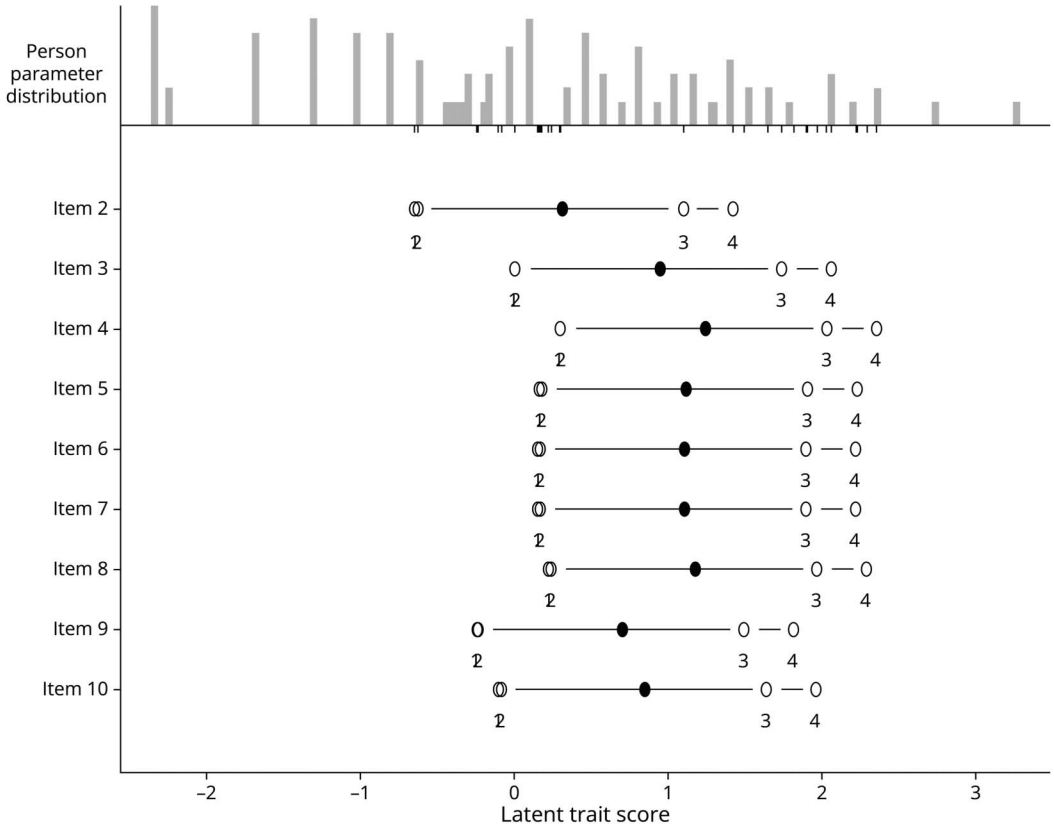
In this study, we present a validated, disease-specific PROM for OMG. The OMGRate-q is predominately focused on diplopia (9 of 10 items) with 1 ptosis item. It has strong psychometric properties on the diplopia subscale. There is a single ptosis item that did not fit well within the overall model. Therefore, future work is needed to develop more items that reflect ptosis so that a robust ptosis subscale can be validated. However, both the diplopia OMGRate-q subscale and the ptosis item were significantly correlated with MGII and MG-QOL-15, suggesting that it is capturing both fluctuation in OMG symptoms and impact of OMG on quality of life, thereby establishing construct validity of the measure. In addition, responses were found to be reliable on repeat testing. The OMGRate-q was acceptable to patients and feasible to use in busy clinical settings as well, taking less than 2 minutes to complete for nearly all participants. Therefore, the OMGRate-q may be used to facilitate future patient-centered research, clinical trials, and clinical care among patients with OMG.

The items included in the OMGRate-q represented a clear diplopia domain and possibly ptosis domain, although the latter needs further development. From a content perspective, both diplopia and ptosis are relevant for vision-related functioning in OMG. Unsurprisingly, the diplopia scale is made of items that relate primarily to symptoms of diplopia and the impact of eye misalignment on visual functioning. 3 of the 4 items with the highest difficulty (highest item locations) were related to diplopia during reading, primary gaze, and downgaze, findings that are corroborated in our previous qualitative

Table 5 Local Item Independence Using the Residual Correlation Matrix

	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 2	1								
Item 3	0.074	1							
Item 4	0.064	0.496	1						
Item 5	−0.109	−0.173	−0.100	1					
Item 6	−0.092	−0.320	−0.342	0.090	1				
Item 7	−0.280	−0.320	−0.362	−0.118	0.149	1			
Item 8	−0.393	−0.361	−0.347	−0.013	0.185	−0.029	1		
Item 9	−0.176	−0.291	−0.297	−0.254	−0.229	0.007	0.142	1	
Item 10	−0.307	−0.298	−0.297	−0.150	−0.163	0.163	0.048	0.029	1

Figure 1 Person-Item Map of Items Used for Latent Trait Estimation



For each item, there is 1 location (closed dot) and there are 4 thresholds (open dots). The first and second thresholds have values that are very close to each other, thus appearing as 1 dot (bolded open dot). Each item has 5 levels ("never," "rarely," "sometimes," "often," and "always") resulting in 4 thresholds. By knowing the latent trait score (x-axis), the highest probability for answering each item with any of the 5 answer options can be determined.

interviews, while the fourth item was related to depth perception. We kept 2 items with identical item locations and threshold parameters (item 6: diplopia looking straight ahead in the distance and item 7: diplopia looking up) because they are capturing diplopia in different gazes. However, future work in larger cohorts may indicate whether one of these items can be reduced. Item threshold plots suggest that respondents seemed to have difficulty distinguishing between adjacent response options (e.g., "never" and "rarely"); future work in larger cohorts will help determine whether collapsing response options is needed. In addition, visual dysfunction from diplopia was not well captured in respondents with very

Table 6 Differential Item Functioning by Comparing Means of Item Residuals

	Age			Site			Gender		
	t	*df*	*p* Value	*t*	*df*	*p* Value	*t*	*df*	*p* Value
Item 2	0.10	130.80	0.92	0.95	131.96	0.34	−0.06	75.54	0.95
Item 3	−1.20	130.18	0.23	0.28	125.43	0.78	0.69	76.42	0.50
Item 4	−1.35	129.99	0.18	0.42	118.60	0.68	−0.60	91.80	0.55
Item 5	−0.39	116.12	0.70	1.63	122.48	0.11	0.75	96.34	0.45
Item 6	0.29	125.97	0.77	0.07	102.75	0.95	0.71	104.30	0.48
Item 7	0.74	130.72	0.46	−0.96	123.42	0.34	1.08	80.33	0.29
Item 8	−0.35	126.13	0.73	−1.44	107.24	0.15	−0.66	108.04	0.51
Item 9	1.21	117.53	0.23	−0.69	113.29	0.49	−1.81	103.44	0.07
Item 10	0.95	119.77	0.35	−0.09	110.34	0.93	0.23	82.52	0.82

mild symptoms—a problem with many MG-related PROMs. This may be due to our population, which had largely stable OMG, or may suggest need for additional items that capture the lowest degree of severity.

We anticipated that 2 items relating to the difficulty with judging distances and depth perception may capture visual dysfunction from ptosis due to loss of binocular vision. On the exploratory factor analysis, the ptosis item (item 1) had the lowest loading value, although it did not clearly indicate that it was a separate factor until significant item misfitting was found in the Rasch model analyses. These findings may be due, in part, to the median ptosis score of 1 indicating that the average respondent rarely had ptosis. Validating the OMGRate-q in a population of patients with more severe disease may be helpful in determining whether difficulty with judging distances and depth perception indeed relate best to other items capturing diplopia. Furthermore, testing additional ptosis-related questions that may capture ptosis severity, fluctuation, or impact on quality of life (e.g., cosmesis) is needed.

We did not observe any DIF based on age or country, which often occurs when groups have different probabilities of response. Likewise, we did not observe any correlation of the latent score with age, sex, seropositivity, or country. Thus, these factors do not seem to have a clear impact on patient-reported visual functioning measured by the OMGRate-q. Furthermore, participants found the OMGRate-q acceptable as assessed with qualitative responses to open-ended questions regarding the survey. We observed that the OMGRate-q value where the disease symptoms likely turn acceptable to not acceptable was at 0.04 (61st percentile), which allows clinicians to determine when additional changes to therapy may be necessary. Future prospective studies should focus on assessing the sensitivity of the OMGRate-q to initiating or changing therapies over time. Similarly, the minimally important difference should be assessed in prospective studies enrolling patients with varying levels of OMG disease severity.

Our study has notable strengths. We followed recommendations for PROM development endorsed by the US Food and Drug Administration, Patient-Reported Outcomes Measurement Information System, and professional bodies.[20] We validated our instrument using a modern test theory approach (polytomous Rasch model) rather than using a classical test theory approach. The strength of this approach is that item's performance can be tested more thoroughly, the measure's precision can be examined with greater detail, and latent trait scores are independent of the items.[21] Furthermore, item response theory offers more flexibility in evaluating patient performance and better investigation of individual item characteristics. We also collected data from multiple centers allowing for greater generalizability to patients with OMG.

To use the OMGRate-q clinically, ptosis is scored by recording the response to a single item. A response to the ptosis item of "often" or "always" indicates that the symptoms are likely unacceptable. However, the diplopia subscale score is calculated from all the remaining items, where the impact of diplopia from OMG is measured with a person's theta, which is estimated from the Rasch model. Theta can be estimated with an iterative algorithm,[12] which we have implemented and provided in an Excel workbook (eAppendix 2). Positive scores indicate visual dysfunction from OMG greater than average in our model-building cohort, which is representative of patients seen at the study sites. Scores range from approximately −3 to +3. An OMGRate-q diplopia subscale score higher than 0.04 (61st percentile) would indicate that the OMG diplopia symptoms are likely *not* acceptable.

However, our study has important limitations. The sample size was relatively small for Rasch analyses, which affected the ability to assess DIF by race. However, our distribution between races was similar to the general US and UK populations, and future studies should enrich cohorts for minority racial groups. We used the Andrich rating scale approach to Rasch analysis for validation, which assumes that all the items have the same discrimination coefficient, unlike other models (e.g., graded response model). However, we selected the approach that best fit our sample size. Our participant population was overall mild and stable, which may have affected our ability to measure visual dysfunction captured in both subscales. In particular, this measure is exclusively aimed at assessing diplopia; therefore, future work is needed to determine how to add measures of ptosis to this scale and whether they can all reflect a single domain of OMG severity or whether a separate ptosis measure is needed. The OMGRate-q also does not have a question about eyelid closure weakness because it is rarely noticed by patients. Last, our study was not designed to directly assess responsiveness of the OMGRate-q to therapies, an important parameter for use in clinical trials and direction for future studies. This is currently planned for the next study.

The OMGRate-q is a novel and valid PROM designed to measure visual dysfunction among patients with OMG. The OMGRate-q emphasizes the impact of diplopia on visual functioning with a single separate item measuring ptosis. Future studies should determine OMGRate-q responsiveness to disease-state changes and how to add measures of ptosis to this scale or whether a separate measure is needed. However, the OMGRate-q addresses an important gap in reliably capturing patient-related visual dysfunction from OMG, and our findings support its clinical and research use.

Acknowledgment

The authors thank the MGFA for funding this study.

Study Funding

This work was funded in part by the Myasthenia Gravis Foundation of America High-Impact Pilot Project Awards (L.B. De Lott and S.H. Wong). L.B. De Lott is supported by an unrestricted institutional grant from the Research to Prevent Blindness and NIH K23EY027849.

Disclosure

S.H. Wong has received research support (paid to her institution) from Visual Snow Initiative, myaware, and MGFA; has received honoraria/consulting fees from argenx and Immunovant; and is the primary developer of the OMGRate and may receive royalties. E. Eggenberger is a co-developer of the OMGRate and may receive royalties. W.T. Cornblath is a co-developer of the OMGRate and may receive royalties. C. Barnett-Tapia has served as member of the advisory board for argenx, Alexion, UCB, and Janssen; has been a consultant for argenx, Janssen, and UCB; has received research support from US Department of Defense, Muscular Dystrophy Canada and MGNet, Grifols, and Octapharma; and is the primary developer of the MGII and may receive royalties. L.B. De Lott is supported by an unrestricted institutional grant from the Research to Prevent Blindness and NIH K23EY027849, and is a co-developer of the OMGRate and may receive royalties. All other authors have no disclosures. Go to Neurology.org/N for full disclosures.

Publication History

Received by *Neurology* June 25, 2024. Accepted in final form October 14, 2024. Submitted and externally peer reviewed. The handling editor was Associate Editor Brian C. Callaghan, MD, MS, FAAN.

Appendix Authors

Name	Location	Contribution
Sui Hsien Wong	Department of Neuro-Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust; Departments of Ophthalmology and Neurology, Guy's and St Thomas' NHS Foundation Trust, London; Department of Ophthalmology, Faculty of Life Sciences & Medicine, King's College London; Department of Clinical and Movement Neuroscience, UCL Queen Square Institute of Neurology, University College London, United Kingdom	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data
Rebhi Abuzaitoun	Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, University of Michigan, Ann Arbor	Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data
Wayne T. Cornblath	Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, University of Michigan, Ann Arbor	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design
Eric R. Eggenberger	Departments of Ophthalmology, Neurology, and Neurosurgery, Mayo Clinic, FL	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design
Sangeeta Khanna	Departments of Ophthalmology, Neurology, and Neurosurgery, Mayo Clinic, FL	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Tatiana Deveney	Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, University of Michigan, Ann Arbor	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data
Joshua R. Ehrlich	Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, and Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI	Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data
Chris A. Andrews	Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, University of Michigan, Ann Arbor	Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data
Carolina Barnett-Tapia	Division of Neurology, Department of Medicine, University Health Network and University of Toronto, Ontario, Canada	Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data
Lindsey B. De Lott	Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, University of Michigan, Ann Arbor	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data

References

1. Saleem A, Wong S. *Quantifying Severity of Ocular Myasthenia Gravis for Research Studies*. North American Neuro-ophthalmology Society Meeting; 2017.
2. Wong SH, Eggenberger E, Cornblath W, et al. Preliminary findings of a dedicated ocular myasthenia gravis rating scale: the OMGRate. *Neuroophthalmology*. 2020; 44(3):148-156. doi:10.1080/01658107.2019.1660686
3. Wong SH, Benatar M. Ocular myasthenia. In: *Myasthenia Gravis and Related Disorders*. H.J. Kaminski & L.L. Kusner. Humana Press; 2018:101-112.
4. Raphael BA, Galetta KM, Jacobs DA, et al. Validation and test characteristics of a 10-item neuro-ophthalmic supplement to the NEI-VFQ-25. *Am J Ophthalmol*. 2006; 142(6):1026-1035. doi:10.1016/j.ajo.2006.06.060
5. Hatt SR, Leske DA, Holmes JM. Comparing methods of quantifying diplopia. *Ophthalmology*. 2007;114(12):2316-2322. doi:10.1016/j.ophtha.2007.01.033
6. Holmes JM, Liebermann L, Hatt SR, Smith SJ, Leske DA. Quantifying diplopia with a questionnaire. *Ophthalmology*. 2013;120(7):1492-1496. doi:10.1016/j.ophtha.2012.12.032
7. Barnett C, Bril V, Kapral M, Kulkarni AV, Davis AM. Myasthenia Gravis Impairment Index: responsiveness, meaningful change, and relative efficiency. *Neurology*. 2017; 89(23):2357-2364. doi:10.1212/WNL.0000000000004676
8. Wood ND, Akloubou Gnonhosou DC, Bowling JW. Combining parallel and exploratory factor analysis in identifying relationship scales in secondary data. *Marriage Fam Rev*. 2015;51(5):385-395. doi:10.1080/01494929.2015.1059785
9. Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. *Psychol Bull*. 1988;103(2):265-275. doi:10.1037/0033-2909.103.2.265

10. Müller M. Item fit statistics for Rasch analysis: can we trust them? *J Stat Distrib Appl*. 2020;7:5. doi:10.1186/s40488-020-00108-7
11. Wright BD, Linacre JM. Reasonable mean-square fit values [online]. rasch.org/rmt/rmt83b.htm.
12. Linacre JM. Estimating Rasch measures with known polytomous (or rating scale) item difficulties anchored maximum likelihood estimation (AMLE). Accessed May 25, 2024. rasch.org/rmt/rmt122q.htm
13. Linacre J. Table 23.99: Largest residual correlations for items: Winsteps Help. Accessed May 25, 2024. https://www.winsteps.com/winman/table23_99.htm
14. Quadri N, Wild D, Skerritt B, Muehlhausen W, O'Donohoe P. A literature review of the variance in interval length between administrations for assessment of test retest reliability and equivalence of pro measures. *Value Health*. 2013;16(3):A40-A41. doi:10.1016/j.jval.2013.03.230
15. Souza MAP, Coster WJ, Mancini MC, Dutra FCMS, Kramer J, Sampaio RF. Rasch analysis of the participation scale (P-scale): usefulness of the P-scale to a rehabilitation services network. *BMC Public Health*. 2017;17:934. doi:10.1186/s12889-017-4945-9
16. Salaffi F, Carotti M, Gutierrez M, Di Carlo M, De Angelis R. Patient acceptable symptom state in self-report questionnaires and composite clinical disease index for assessing rheumatoid arthritis activity: identification of cut-off points for routine care. *Biomed Res Int*. 2015;2015:930756. doi:10.1155/2015/930756.
17. Mendoza M, Tran C, Bril V, Katzberg HD, Barnett C. Patient-acceptable symptom states in myasthenia gravis. *Neurology*. 2020;95(12):e1617-e1628. doi:10.1212/WNL.0000000000010574
18. Mair P, Hatzinger R. Extended Rasch modeling: the eRm package for the application of IRT models in R. *J Stat Softw*. 2007;20(9):1-20. doi:10.18637/jss.v020.i09
19. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021.
20. US Department of Health and Human Services FDA Center for Drug Evaluation and Research, US Department of Health and Human Services FDA Center for Biologics Evaluation and Research, US Department of Health and Human Services FDA Center for Devices and Radiological Health, , , et al. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*. 2006;4:79. doi:10.1186/1477-7525-4-79
21. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. *Patient*. 2014;7(1):23-35. doi:10.1007/s40271-013-0041-0