



Validity and reliability of the Persian version of the myasthenia gravis impairment index

Received: 08 Sep. 2025
Accepted: 05 Nov. 2025

Seyed Jalaeddin Hadei^{1,2}, Payam Khomand³, Farzad Fatehi^{1,4}, Vera Bril⁵, Carolina Barnett⁵, Shahriar Nafissi^{1,2}

¹ Neuromuscular Research Center, Tehran University of Medical Sciences, Tehran, Iran

² Department of Neurology, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran

³ Dalhousie University, Department of Neurology, Saint John Regional Hospital, Saint John, NB, Canada

⁴ Neurology Department, University Hospital of Leicester NHS Trust, Leicester, UK

⁵ Division of Neurology, Department of Medicine, University Health Network, Toronto, Canada

Keywords

Cross-Cultural Care; Reliability and Validity; Responsiveness; Myasthenia Gravis

Abstract

Background: The Myasthenia Gravis Impairment Index (MGII) is a valid and reliable tool for assessing patients with myasthenia gravis (MG), originally developed in English. This study evaluates the validity, reliability, and responsiveness of a Persian translation of the MGII in Iran.

Methods: The MGII was translated through a multi-step process for cultural adaptation. Content validity was assessed via face validity, Content Validity Index (CVI), and Content Validity Ratio (CVR). Construct validity was evaluated by correlating MGII scores with the Myasthenia Gravis Composite (MGC), Quantitative Myasthenia Gravis (QMG) score, Myasthenia Gravis

Activities of Daily Living Scale (MG-ADL), and 15-item Myasthenia Gravis Quality of Life (MGQOL-15). Test-retest reliability was assessed using the Intraclass Correlation Coefficient (ICC) in stable patients. Additionally, responsiveness was evaluated by comparing pre- and post-treatment MGII scores and their correlation with other measures.

Results: One hundred patients with MG participated in the study. The Persian version of the MGII demonstrated good face and content validity, excellent test-retest reliability (ICC > 0.90 for most scores), and good inter-rater reliability (ICC > 0.80). The floor effect was low at 4%.

How to cite this article: Hadei SJ, Khomand P, Fatehi F, Bril V, Barnett C, Nafissi S. **Validity and reliability of the Persian version of the myasthenia gravis impairment index.** Curr J Neurol 2026; 25(1): 55-64.

A significant improvement in total MGII score (-18.9 ± 12.6) was noted in 26 patients after treatment changes, with moderate to high correlations observed between changes in the MGII scores and changes in the MGC, QMG, MG-ADL, and MGQOL-15 scores.

Conclusion: This study exhibits the validity and reliability of the Persian version of MGII, supporting its use in assessing patients with MG and their response to treatment in the Iranian population.

Introduction

Myasthenia gravis (MG) is an autoimmune disease characterized by fluctuating weakness of the voluntary muscles. Muscle weakness varies widely in severity, ranging from mild involvement of the extraocular muscles to severe weakness affecting the respiratory, bulbar, and limb muscles. The underlying pathology involves autoantibodies targeting the neuromuscular junction, and the specific antibody profile can influence disease severity and clinical features. MG can significantly impact a patient's quality of life (QoL) and even lead to mortality.¹⁻³

Accurate clinical assessment is essential for effective management in MG. However, the fluctuating nature of the disease presents challenges.^{4,5} Several tools exist to assess MG severity, including the Quantitative Myasthenia Gravis (QMG) score, the Myasthenia Gravis Composite (MGC) score, the Myasthenia Gravis Activities of Daily Living (MG-ADL), and the 15-item Myasthenia Gravis Quality of Life (MGQOL-15). Each tool offers unique advantages and limitations, with some relying on expert opinion and others on patient-reported data.⁶⁻⁹

The Myasthenia Gravis Impairment Index (MGII) is a new measure of MG severity; its feasibility, reliability, and construct validity were established by Barnett et al. in 2016. The MGII has been developed based on the guidelines provided by the United States Food and Drug Administration (FDA) and Consensus-based Standards for the selection of health Measurement Instruments (COSMIN), incorporating patient input. This method has been found to have a lower floor effect compared to other measures. The MGII comprises 22 patient-reported items and 6 examination-based items, with total score ranging from 0 to 84; higher scores indicate more severe impairment. The index is divided into two subscales: ocular, which assesses eight items, and generalized, which evaluates 20 items related to other body areas.¹⁰

Importantly, the MGII has shown sensitivity in detecting treatment response. Specifically, patients receiving therapies such as prednisone, intravenous immunoglobulin (IVIG), or plasma exchange (PLEX) exhibited significantly greater improvements in MGII scores compared to controls.¹¹ Due to floor effects, some previous measures may not show significant responses to the treatments. The index, initially developed and validated in Canada, recently underwent validity assessment for its Dutch and Italian versions.^{10,12,13}

Further cross-cultural validation is necessary to confirm the MGII applicability in diverse populations. This study aimed to address the gap by translating and culturally adapting the MGII for Persian-speaking patients with MG. Subsequently, we evaluated the validity, reliability, and responsiveness of the Persian version in this specific population.

Materials and Methods

This study was conducted at the Neurology Department of Shariati Hospital, Tehran, Iran, between August 2022 and September 2023. The Study Protocol was approved by the Ethics Committee of Tehran University of Medical Sciences (Approval ID: IR.TUMS.MEDICINE.REC.1401.136). Written consent was obtained from all participants.

Participants

A total of 100 participants were enrolled in this study. Participants were consecutive patients with MG at different disease stages who attended the outpatient clinic or were hospitalized in the neurology ward. Exclusion criteria comprised MG crisis or intubation, which implies class V on the Myasthenia Gravis Foundation of America (MGFA) Clinical Classification.¹⁴ Furthermore, individuals with chronic conditions, such as diabetes, cardiovascular or cerebrovascular events, asthma, and renal or hepatic dysfunction, were excluded to avoid confounding the health-related QoL (HRQoL) measurements. Additionally, those with major psychiatric conditions, such as major depression or psychosis, were also excluded.¹⁵ Patients were diagnosed with MG based on their fatigability and fluctuating weakness, along with paraclinical confirmatory evaluations, including an increased plasma level of acetylcholine receptor (AChR) or muscle-specific tyrosine kinase (MuSK) antibodies, decremental responses on low-frequency repetitive nerve stimulation, or increased jitter on single-fiber electromyography (EMG).

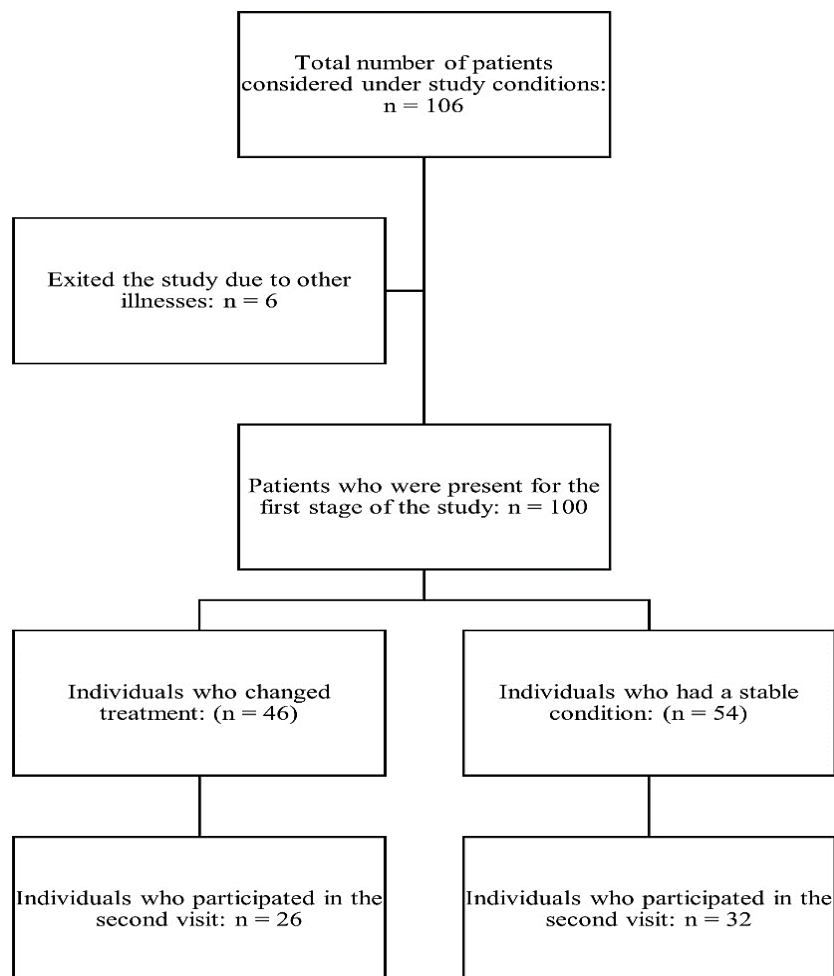


Figure 1. Study flowchart

A positive Tensilon test also contributed to the diagnosis. All patients were receiving standard MG treatment, such as anticholinesterase, prednisolone, immunosuppressive agents, IVIG, or PLEX.

During the first visit, patients who had not undergone any treatment changes for at least the previous 3 months were asked to complete the Persian version of MGQOL-15 and the 22-item self-report part of MGII. The remaining six items of the examination¹⁰ were performed by a neurologist as the first rater. After an hour, another blind examiner assessed the examination items of the MGII in 32 patients to ensure inter-rater validity. We evaluated the severity of the MG using the QMG, MGC, MG-ADL, and MGFA. According to the definitions provided by the Task Force of the Medical/Scientific Advisory Board of the MGFA, patients were classified into ocular and generalized groups. Patients with generalized MG were further subdivided into mild (Class II), moderate (Class III), and severe (Class IV) categories.¹⁴ Second consecutive visits were

performed 4–8 weeks after the first visit in both stable patients and those undergoing treatment modifications (including the start or dose adjustment of prednisolone, other cytotoxic medications, IVIG, or PLEX), as clinical improvement is expected to be temporary over this period.^{16,17} The appropriate dosage for each patient's medication was determined by a neurologist or neuromuscular specialist. The patients were administered IVIG or PLEX according to standardized protocols.¹⁷

Translation

To adapt the MGII from English to Persian, we obtained permission from the corresponding author of the original MGII article¹⁰ and applied cross-cultural adaptation guidelines, considering cultural adjustment and validation of the Persian version of the MGII.^{18–20}

A conventional multi-step forward-backward technique was applied to convert the MGII questionnaire to Persian. A physician at our center translated the original English version of the MGII

into Persian. After that, another independent physician, blinded to the primary assessment, back-translated this Persian version to English. The back-translation was then balanced to the earliest, and components with significant dissimilarities were modified.²⁰

A committee composed of two translators, four neurologists, and two experts in linguistics evaluated the original MGII questionnaire, the forward translations, the back translations, and the discussed incompatibilities.

Outcome Measurements

MGII: The MGII assesses the severity of MG symptoms using 28 items on a scale ranging from 0 to 84, where higher scores indicate greater clinical severity. The scale comprises two subscales: ocular (8 items) and generalized (20 items). The patient-reported component consists of 22 items encompassing diplopia (3), ptosis (3), swallowing (1), chewing (2), voice (3), articulation (3), breathing (1), general fatigue (1), neck weakness (1), and upper and lower limb weakness (2 each). Additionally, six examination items evaluate diplopia, ptosis, lower facial strength, neck stamina, and upper and lower limb strength.¹⁰

MGQOL-15: The MGQOL-15 is a self-administered questionnaire that assesses the QoL in patients with MG. It consists of 15 items that cover various aspects of daily life. Each item is rated on a scale of 0 to 4, with 0 representing no problems and 4 representing severe problems. The total score ranges from 0 to 60, with higher scores indicating a lower QoL.¹⁵

MGC: The MGC is a 10-item measure that assesses the severity of MG symptoms. It includes items on muscle weakness, fatigue, ptosis, diplopia, dysphagia, dysarthria, and limb weakness. The total score ranges from 0 to 50, with higher scores indicating more severe symptoms.⁸

QMG: The QMG is a 13-item measure that quantitatively assesses MG severity. It consists of the following domains: ocular (2 items), facial (1), bulbar (2), gross motor (6), axial (1), and ocular (1). Each item is scored from 0 to 3 (with 3 being the most severe), giving an overall QMG score from 0 to 39. The higher scores denote increased disease severity.²¹

MG-ADL: The MG-ADL is an 8-item questionnaire that assesses the impact of MG on daily activities. Each item is rated on a scale of 0 to 3, with 0 representing no difficulty and 3 representing extreme difficulty. The total score ranges from 0 to 24, with higher scores indicating

greater disability.²¹

MGFA: The MGFA classification is a clinical staging system for MG that categorizes patients into five classes based on the severity of their symptoms. Class I patients have only ocular symptoms, while Class V patients are ventilator-dependent due to myasthenic crisis.¹⁴

Study Size

Recruiting large samples for research studies is challenging because of the rarity of MG. Furthermore, there is no consensus regarding the optimal sample size for studies evaluating the measurement properties of an instrument. For construct validity analyses, a minimum sample size of 62 participants is required to detect a correlation coefficient of 0.40, with a significance level of 0.05 and a statistical power of 90%. Moreover, according to the COSMIN framework, a sample size of 100 participants is considered to represent excellent methodological quality. Therefore, in accordance with COSMIN recommendations, we aimed to recruit 100 participants.¹³

Analysis was conducted using SPSS software (version 27, IBM Corporation, Armonk, NY, USA), and a P-value of 0.05 was considered statistically significant. The compliance of variables with a normal distribution was tested using the Shapiro-Wilk test and probability graphics. Results were presented as median and interquartile range (IQR) for non-normally distributed data, whereas data with normal distributions were provided as mean $\pm$ standard deviation (SD).

Reliability

Inter-rater reliability was assessed on the first visit day, with an approximately one-hour interval between the two blinded raters. Weighted kappa coefficients were computed to quantify agreement.²² During the second consecutive visit, 32 patients who reported no changes in their symptoms or medications over a period of 4 to 8 weeks were evaluated for test-retest reliability. Using a random effects model, the Intraclass Correlation Coefficient (ICC_{2,1}) was calculated for the total score as well as for the subscales. ICC values were reported with 95% confidence intervals (CIs). Within this range, ICC values are recommended at least 0.80 for groups and > 0.90 for individuals. Generally, values between 0.60 and 0.80 are considered substantial, while 0.80 is considered excellent agreement. However, there is no universal consensus on the interpretation of kappa.^{13,23} Furthermore, the standard error of measurement (SEM) was calculated for the total MGII score.

Validity

To assess the face validity of the Persian MGII, 10 experts in MG were consulted. Subsequently, content validity was evaluated using the Content Validity Ratio (CVR) and Content Validity Index (CVI) obtained from the neurologist's responses about different aspects of the Persian version of MGII.^{24,25} Construct validity was evaluated by assessing the correlations between the Persian MGII scores with scores on the Persian versions of QMG, MGC, MG-ADL, and MGQOL-15 using Pearson correlation coefficients. It was hypothesized that these correlations would align with the ranges observed in the original validation of the MGII ($r = 0.60-0.80$).¹⁰

Responsiveness

Patients who received IVIG or PLEX, or whose prednisolone dose was increased by more than 50% relative to the preceding dose, were included in the responsiveness assessment of the Persian version of the MGII. At follow-up, patients were clinically categorized into four post-treatment states: improved, unchanged, worsened, or exacerbated, based on the overall assessment of disease status. Scores on the MGII, QMG, MGC, MG-ADL, and MGQOL-15 before and after treatment were compared.

Paired t-tests and analysis of covariance (ANCOVA) were used to assess changes in scores, adjusting for baseline values. Additionally, responsiveness was assessed utilizing the standardized response mean (SRM) and effect size (ES). A dependable anchor-based minimal clinically important difference (MCID) was not possible due to the anchor's inability to differentiate between minimal and marked improvement. As a result, the MCID was estimated using a distribution-based method corresponding to one-half of the baseline SD. Furthermore, the minimal detectable change at the 95% confidence level (MDC95) was calculated.

Results

Patient Characteristics and Clinical Features

During the initial visit, 106 patients with MG were recruited, of which 100 were finally included in the study analysis. The mean age was 45.3 ± 14.1 years, with 57% being women. As shown in table 1, the median disease and treatment durations at baseline were 44 months (IQR: 12-123) and 36 months (IQR: 5-120), respectively. Prednisolone and pyridostigmine were administered to 84% and 81% of patients. Thymectomy was performed in

21% of patients, and 9% had a thymoma. Patients were divided into two groups: stable ($n = 54$) and exacerbation/worsening ($n = 46$). Disease duration was significantly longer in the stable group than in the exacerbation/worsening group (84 months vs. 24 months; $P < 0.001$). Similarly, treatment duration was also higher in the stable group (77 months vs. 47 months; $P < 0.001$). Additionally, there was a significant association between disease stability and other factors, including residence ($P = 0.040$) and azathioprine use ($P = 0.005$).

Face Validity

The face validity of the Persian MGII was evaluated by a panel of ten experts in MG. Their feedback was collected and analyzed, and the experts generally agreed that the questionnaire demonstrated good face validity, noting that the items were clear, understandable, and directly relevant to measuring MG symptoms. Additionally, the sequencing of the items was deemed logical and appropriate. However, some experts suggested refinements for improved clarity. Based on their feedback, specific items were rephrased for better understanding. For example, the item on lower extremity fatigue during activity originally referenced the block test for walking distance. Considering cultural and geographical variations (urban vs. rural settings), the item was modified to assess the ability to walk 1000 meters instead of a specific number of blocks.

Content Validity

Ten neurology residents participated in the assessment of the Persian MGII's content validity. All items achieved a CVR greater than 0.62, indicating an acceptable level of relevance to the intended construct (MG symptoms). Furthermore, the CVI of each item and the overall questionnaire surpassed the recommended threshold of 0.79, providing further evidence for the instrument's content validity.

Construct Validity

The most effective way to assess construct validity is by comparing it to a gold standard. In the absence of a gold standard for evaluating symptom severity and clinical status in patients with MG, other validated tools that are reliable in Persian and measure similar constructs were used. Mean MGII total scores escalated steadily with increasing MGFA clinical classes. Patients in remission exhibited a mean MGII score of 0.8, compared to 10.7 for class I, 13.7 for class IIa, 14.1 for class IIb, 27.2 for class IIIa, 36.6 for class IIIb, 52.7 for class IVa, and 58.7 for class IVb.

Table 1. Descriptive statistics of demographic and clinical variables at baseline

Variable	Category	Total (n = 100)	Status		P
			Stable (n = 54)	Worsening (n = 46)	
Age (year)	-	45.3 ± 14.8	45.3 ± 13.0	45.3 ± 16.8	0.900
Duration of illness (month)	-	44.0 (12.0–123.0)	84.0 (36.0–177.0)	24.0 (5.0–72.0)	< 0.001
Duration of treatment (month)	-	36.0 (5.0–120.0)	77.0 (25.2–154.5)	47.0 (0.5–69.0)	< 0.001
Gender	Women	57 (57.0)	34 (62.9)	23 (50.0)	0.200
	Men	43 (43.0)	20 (37.0)	23 (50.0)	
Marital status	Married	79 (79.0)	45 (83.3)	34 (73.9)	0.200
	Single	20 (20.0)	8 (14.8)	12 (26.0)	
	Divorced	1 (1.0)	1 (1.8)	0 (0)	0.800
Employment	Employed	19 (19.0)	9 (16.6)	10 (21.7)	
	Unemployed	70 (70.0)	40 (74.0)	30 (65.2)	
	Retired	9 (9.0)	4 (7.4)	5 (10.8)	
	Disabled	2 (2.0)	1 (1.8)	1 (2.1)	0.040
Residence	Capital city (Tehran)	66 (66.0)	41 (75.9)	25 (54.3)	
	Other cities	33 (33.0)	13 (24.0)	20 (43.4)	
	Rural areas	1 (1.0)	0 (0)	1 (2.1)	0.060
Treatment status	Untreated	12 (12.0)	3 (5.5)	9 (19.5)	
Medication use	Prednisolone	84 (84.0)	48 (88.8)	36 (78.2)	0.100
	Pyridostigmine	81 (81.0)	44 (81.4)	37 (80.4)	> 0.999
	Azathioprine	57 (57.0)	38 (70.3)	19 (41.3)	0.005
	Rituximab	7 (7.0)	3 (5.5)	4 (8.7)	0.700
	Mycophenolate	2 (2.0)	2 (3.7)	0 (0)	0.400
Surgical history	Thymectomy	21 (21.0)	12 (22.2)	9 (19.5)	0.800
Clinical characteristics	Thymoma	9 (9.0)	5 (9.2)	4 (8.7)	> 0.999
Antibody status	AChR	73 (73.0)	37 (68.5)	36 (78.2)	0.300
	MuSK	8 (8.0)	6 (11.1)	2 (4.3)	
	Seronegative	19 (19.0)	11 (20.3)	8 (17.3)	0.700
Type of MG	Generalized	91 (91.0)	50 (92.5)	41 (89.1)	
	Ocular	9 (9.0)	4 (7.4)	5 (10.8)	N/A
MGFA classification*	0	5 (5.0)	5 (9.3)	0 (0)	
	I	9 (9.0)	5 (9.3)	4 (8.7)	
	IIa	17 (17.0)	16 (29.6)	1 (2.2)	
	IIb	10 (10.0)	9 (16.7)	1 (2.2)	
	IIIa	17 (17.0)	12 (22.2)	5 (10.9)	
	IIIb	27 (27.0)	7 (13.0)	20 (43.5)	
	IVa	4 (4.0)	0 (0)	4 (8.7)	
	IVb	11 (11.0)	0 (0)	11 (23.9)	

Normally distributed quantitative variables are presented as mean ± standard deviation (SD), while non-normally distributed ones are reported as median and interquartile range (IQR). Qualitative variables are described by frequency (percent). Independent t-tests were used to assess group differences, while Mann–Whitney U tests were employed to analyze non-normally distributed variables. Associations between qualitative variables and status were evaluated using Fisher's exact test.

*MGFA classes: remission = no symptoms for > 1 year; I = Pure ocular; II = Mild generalized; III = Moderate generalized; IV = Severe generalized

AChR: Acetylcholine receptor; MG: Myasthenia gravis; MGFA: Myasthenia Gravis Foundation of America; MuSK: Muscle-specific kinase; N/A: Not available

This trend illustrates the MGII's capacity to differentiate among various degrees of disease severity as delineated by the MGFA classification. Table 2 demonstrates that the MGII exhibited strong positive correlations with recognized

metrics of MG severity and QoL, affirming its convergent validity. The most significant correlations were found with the MG-ADL ($r = 0.92$, 95% CI: 0.88–0.94, $P < 0.001$) and MGC ($r = 0.91$, 95% CI: 0.88–0.94, $P < 0.001$), followed by

the QMG ($r = 0.88$, 95% CI: 0.83–0.92, $P < 0.001$) and MGQOL-15 ($r = 0.82$, 95% CI: 0.75–0.87, $P < 0.001$). All observed correlations fulfilled our a priori hypotheses by falling within or exceeding the predefined expected range ($r = 0.60$ – 0.80), and all were statistically significant, supporting the excellent convergent validity of the Persian version of the MGII.

Table 2. Myasthenia Gravis Impairment Index (MGII) correlations with other instruments

Comparison measure	Pearson r (95% CI)	P
QMG	0.88 (0.83–0.92)	< 0.001
MGC	0.91 (0.88–0.94)	< 0.001
MG-ADL	0.92 (0.88–0.94)	< 0.001
MGQOL-15	0.82 (0.75–0.87)	< 0.001

CI: Confidence interval; MGII: Myasthenia Gravis Impairment Index; MGC: Myasthenia Gravis Composite; MG-ADL: Myasthenia Gravis Activities of Daily Living; MGQOL-15: Myasthenia Gravis Quality of Life 15-item; QMG: Quantitative Myasthenia Gravis

Reliability

ICC was calculated for the Persian MGII to determine the test-retest reliability (Table 3). All ICC values for the ocular, generalized, and self-reported items, as well as for the total score, exceeded 0.90, indicating excellent test-retest reliability. This signifies that scores on these dimensions remain consistent across time. Although slightly lower, the ICC values for the examination items remained acceptable (> 0.80). Additionally, the inter-rater reliability for these items yielded an ICC of 0.85.

Responsiveness

To evaluate the responsiveness of the Persian MGII to clinical changes following therapeutic interventions, variations across all dimensions were assessed pre- and post-treatment. Table 4 demonstrates significant improvements across all dimensions in the treatment group ($n = 26$) compared to the control group ($n = 32$) following

treatment ($P < 0.001$). Notably, the treatment group showed substantial reductions in total MGII score (-18.9 ± 12.6), total ocular items score (-7.5 ± 6.9), and total generalized items score (-11.8 ± 8.6) compared to the control group. Table 5 details the changes observed within the treatment group across different dimensions ($P < 0.001$). Patients who underwent treatment modifications exhibited significant improvements in total MGII score (-18.92 ± 12.62), MGC, QMG, MG-ADL, and MGQOL-15 over 4–8 weeks. A distribution-based estimate, corresponding to one-half of the baseline SD, yielded an MCID of approximately 9 points. The SRM and ES were 1.50 and 1.06, respectively, indicating excellent responsiveness. The SEM was 3.11, resulting in an MDC95 of 8.62 points. Collectively, these findings provide further evidence that the MGII is responsive to changes in patients' clinical status.

Discussion

This study aimed to evaluate the validity and reliability of the Persian version of the MGII among Persian-speaking individuals in Iran. Given that responsiveness—the ability of a clinical assessment tool to detect changes over time—is crucial for its practical application in clinical settings and trials, the responsiveness of the Persian version of the MGII was also examined. The results demonstrated good construct, face, and content validity, as well as inter-rater and test-retest reliability of the Persian version of MGII. Moreover, the instrument proved responsive to changes in patients' clinical status.

There are similarities and differences between our patients' characteristics and those reported in previous studies. In our study, 73% of patients had antibodies against AChR and 7% against MuSK, which is roughly similar to Dresser et al.'s report.³ Furthermore, 91% of patients presented with generalized MG, whereas 9% exhibited ocular MG.

Table 3. Reliability of dimensions of the Persian version of Myasthenia Gravis Impairment Index (MGII)

Dimension	N	Test 1	Test 2	SEM	ICC (95% CI)
Ocular items	32	1.5 $\pm$ 4.5	2.0 $\pm$ 4.0	1.42	0.90 (0.82–0.95)
Generalized items	32	10.7 $\pm$ 10.7	9.9 $\pm$ 10.0	2.67	0.93 (0.87–0.96)
Self-report items	32	9.3 $\pm$ 9.1	8.9 $\pm$ 8.9	2.94	0.90 (0.80–0.95)
Examination items (test-retest)	32	3.1 $\pm$ 2.8	3.0 $\pm$ 2.8	1.01	0.87 (0.75–0.93)
Examination items (inter-rater)*	30	3.1 $\pm$ 2.8	2.7 $\pm$ 3.0	1.08	0.85 (0.71–0.92)
Total score	32	12.3 $\pm$ 11.0	12.0 $\pm$ 11.2	3.11	0.92 (0.84–0.96)

Data are presented as mean $\pm$ standard deviation (SD).

*For inter-rater reliability, scores are for rater 1 and rater 2.

CI: Confidence interval; ICC: Interclass correlation coefficient; SEM: Standard error of measurement

Table 4. Treatment response by dimension on the Persian version of Myasthenia Gravis Impairment Index (MGII)

Dimension	Control group (n = 32)	Treatment change group (n = 26)	P
Total MGII	0.2 ± 4.3	-18.9 ± 12.6	0.001
Total ocular items score	-0.5 ± 1.8	-7.5 ± 6.9	0.001
Total generalized items score	0.7 ± 3.5	-11.8 ± 8.6	0.001

Data are presented as mean ± standard deviation (SD).

MGII: Myasthenia Gravis Impairment Index

Table 5. Changes in questionnaire scores in response to treatment modification

Dimension	Mean difference	SD	P
Total MGII	-18.92	12.62	0.001
MGC	-8.07	7.21	0.001
QMG	-5.30	4.89	0.001
MG-ADL	-5.65	4.09	0.001
MGQOL-15	-13.19	10.37	0.001

Changes were measured over a 4–8-week period.

MGII: Myasthenia Gravis Impairment Index; MGC: Myasthenia Gravis Composite; MG-ADL: Myasthenia Gravis Activities of Daily Living; MGQOL-15: Myasthenia Gravis Quality of Life 15-item; QMG: Quantitative Myasthenia Gravis; SD: Standard deviation

These values are similar to previous studies reporting that generalized MG accounts for approximately 80%–85% of cases.³ In addition, 19% of patients were seronegative, slightly exceeding the reported value in the literature (10%–15%), but closely matching the findings of the Dutch study, in which 20 out of 99 patients were seronegative.^{2,12} Since our hospital is a referral center for neuromuscular disorders, a higher proportion of MG patients with atypical findings, such as seronegativity, may be expected compared with the general MG population. Moreover, the higher prevalence of seronegativity among our patients may result from prior treatment effects. Nine percent of our patients had thymoma, which aligns closely with earlier reports.

The MGII was translated using the backward-forward method. Consistent with the Italian and Dutch versions,^{12,13} and considering differences in urban planning and street layouts, item 21 which assesses lower limb weakness during walking was adapted by expressing distance in meters rather than city blocks.

The results of this study indicated that higher disease severity, as categorized by the MGFA classification, was associated with higher total MGII scores, providing evidence of construct validity. To further evaluate construct validity, Pearson's correlation coefficients were calculated to evaluate the relationships between the MGII and other tools mentioned. The correlation between the MGII and each of the four comparator tools exceeded 0.80, indicating a strong relationship. In

the original development study of the MGII, the correlation values with the specified measures—the MGC, QMG, MG-ADL, and MGQOL-15—were 0.81, 0.79, 0.91, and 0.79, respectively. These findings are similar to our study and demonstrate a high level of correlation.¹⁰ The results of the Italian study, which used Spearman's rank correlation coefficient, and the Dutch study were consistent with those of the present study and the original version, indicating high correlations (all $r > 0.70$, except $r = 0.68$ for QMG in the Dutch study).^{12,13}

In the present study, the ICC value for the total MGII score was 0.92 (0.84–0.96), indicating very high test-retest reliability and good repeatability of the MGII across repeated assessments. In the item analysis, the ICCs for ocular items, generalized items, and patient self-reported items all exceeded 0.90, indicating the very high reliability of these sections of the questionnaire. The ICC for the examination section in the test-retest method was 0.87, and the inter-rater ICC was 0.85, both of which denote good reliability for this section of the instrument. In the original version of the questionnaire, the ICC values for the total score, ocular items, generalized items, self-reported items, and examination items in the test-retest assessment were 0.92, 0.87, 0.92, 0.91, and 0.90, respectively, with an inter-rater ICC of 0.80.¹⁰ In the Italian version, which included 25 clinically stable patients at the follow-up visit, the ICCs for the total MGII score and the self-reported subscale were 0.93 and 0.91, respectively. Additionally, the inter-rater ICC for the examination section was reported as 0.74, indicating acceptable reliability.¹³ The results of these two studies are consistent with those of the present study, emphasizing the reliability of MGII across different populations.

The floor effect for the total MGII score in this study was 4%, which was lower than that observed for the MGC, MG-ADL, and MGQOL-15 (6%, 13%, and 7%, respectively), but higher than that for QMG (1%). The lower floor effect in QMG could be attributed to the longer time required to assess limb and neck strength compared with similar items in the MGII. However, in the Dutch version of the study by De Meel et al., the floor effect of

QMG was higher than that of MGII.¹² In the Italian study and the original version, the floor effect of QMG was not mentioned. In line with the Canadian and Italian studies, the largest floor effect was for MG-ADL in this study.^{10,13}

The primary objective of the original design of the MGII was to enhance the monitoring of patients with MG post-treatment.^{10,11} It is crucial to estimate the actual change in the progression of the disease rather than focusing on the SEM of the instrument. Using a distribution-based approach, the MCID of the Persian MGII was estimated at approximately 9 points. Although this estimate cannot be directly compared with the anchor-based values reported by Barnett et al. (5.5 points for individuals and 8.1 points for groups), it is close to their group-level minimal important difference (MID). Likewise, the MDC95 of the Persian MGII was 8.62 points, closely approximating the 9.1 points reported by Barnett et al. for the original English version, further supporting the MGII's reproducibility across different populations.¹¹

Responsiveness to treatment in MGII was assessed in 26 patients with worsening symptoms. Following treatment, 14 patients showed clear improvement, 9 demonstrated partial improvement, and 3 reported no change. Despite the absence of perceived improvement in 3 patients, the mean MGII score (5.8) and scores on other assessment measures showed clear decreases compared to baseline. Since it takes an average of two weeks for prednisolone to start working and up to six months for its maximum effect,²⁶ the greater decrease in the mean MGII score in our patients (18 points) compared to Barnett et al.'s study (7 points) can be attributed to the later return of our patients (4–8 weeks compared to 3–4 weeks) and the greater impact of treatment.¹¹ Among patients whose clinical status changed following treatment, mean MGC, QMG, MG-ADL, and MGQOL-15 scores decreased by approximately 8.0, 5.3, 5.6, and 13.1 points, respectively, indicating statistically significant improvements. It is noteworthy that these decreases were greater than those observed in the original validation study by Barnett et al.¹¹ The correlation analysis revealed a strong relationship between changes in the total MGII score and MG-ADL, with moderate correlations observed across other assessments. The MGII changes had the lowest correlation with MGQOL-15. The pure objectivity of MGQOL-15, rather than other measurements and MGII, explains this lowest correlation. In Pasqualin et al.'s study on the Italian

version, a moderate correlation was observed between changes in MGII and MG-ADL, as well as MGC. As in our study, the lowest correlation was reported with the MGQOL-15.¹³

This study is subject to several limitations. First, this was a single-center study conducted at a tertiary referral facility, excluding patients with comorbidities that could affect MG symptoms, consistent with previous MGII validation studies. Therefore, our cohort may not fully represent the broader Iranian MG population, and referral bias may limit the generalizability of the findings. Second, the relatively small sample sizes for the test-retest and responsiveness analyses may have limited the precision of these estimates. Despite consecutive enrollment, larger cohorts were difficult to achieve owing to the rarity of MG and challenges in following up patients from distant regions. Third, while responsiveness evaluation had been scheduled at four-week intervals, postponed follow-up visits for certain participants, attributed to extensive travel distances and the high patient volume at our public referral facility, led to inconsistent assessment intervals. Finally, divergent construct validity and receiver operating characteristic (ROC) analyses were not performed.

Conclusion

The Persian version of the MGII demonstrated excellent to good reliability and strong construct validity. The presence of patient-reported items alongside examination items, the low floor effect, high responsiveness to changes in patients' clinical status, and good correlation with other outcome measures make this index a robust instrument for assessing the status of patients with MG in stable and clinically exacerbated conditions, as well as in clinical trials.

Acknowledgments

We would like to express our sincere gratitude to all the patients who participated in this study; their willingness to share their experiences has been invaluable to our research. We also extend our thanks to all neurology residents of Shariati Hospital for their assistance in completing the examination items of the MGII. Their dedication and support were crucial to the accuracy and thoroughness of our work. Additionally, we are grateful to all the neurologists and translation team for their meticulous work ensuring that our findings are accessible to a broader audience. Their efforts have been instrumental in bridging

language barriers and enhancing the impact of our research. We thank all contributors for their

essential support.

References

- Oosterhuis HJ. The natural course of myasthenia gravis: a long term follow up study. *J Neurol Neurosurg Psychiatry* 1989; 52(10): 1121-7.
- Lazaridis K, Tzartos SJ. Autoantibody Specificities in Myasthenia Gravis; Implications for Improved Diagnostics and Therapeutics. *Front Immunol* 2020; 11: 212.
- Dresser L, Wlodarski R, Rezaei K, Soliven B. Myasthenia Gravis: Epidemiology, Pathophysiology and Clinical Manifestations. *J Clin Med* 2021; 10(11): 2235.
- Regnault A, Morel T, De la Loge C, Mazerolle F, Kaminski HJ, Habib AA. Measuring Overall Severity of Myasthenia Gravis (MG): Evidence for the Added Value of the MG Symptoms PRO. *Neurol Ther* 2023; 12(5): 1573-90.
- Nafissi S, Okhovat AA, Sinaei F, Ansari B, Ayramloo H, Basiri K, et al. Iranian Consensus Recommendations for Treatment of Myasthenia Gravis. *Arch Iran Med* 2022; 25(1): 37-49.
- Wolfe GI, Herbelin L, Nations SP, Foster B, Bryan WW, Barohn RJ. Myasthenia gravis activities of daily living profile. *Neurology* 1999; 52(7): 1487-9.
- Bedlack RS, Sanders DB. On the concept of myasthenic crisis. *J Clin Neuromuscul Dis* 2002; 4(1): 40-2.
- Burns TM. History of outcome measures for myasthenia gravis. *Muscle Nerve* 2010; 42(1): 5-13.
- Mullins LL, Carpentier MY, Paul RH, Sanders DB. Disease-specific measure of quality of life for myasthenia gravis. *Muscle Nerve* 2008; 38(2): 947-56.
- Barnett C, Bril V, Kapral M, Kulkarni A, Davis AM. Development and validation of the Myasthenia Gravis Impairment Index. *Neurology* 2016; 87(9): 879-86.
- Barnett C, Bril V, Kapral M, Kulkarni AV, Davis AM. Myasthenia Gravis Impairment Index: Responsiveness, meaningful change, and relative efficiency. *Neurology* 2017; 89(23): 2357-64.
- De Meel RHP, Barnett C, Bril V, Tannemaat MR, Verschuuren J. Myasthenia Gravis Impairment Index: Sensitivity for Change in Generalized Muscle Weakness. *J Neuromuscul Dis* 2020; 7(3): 297-300.
- Pasqualin F, Barnett C, Guidoni SV, Albertini E, Ermani M, Bonifati DM. Validation of the Italian version of the Myasthenia Gravis Impairment Index (MGII). *Neurol Sci* 2022; 43(3): 2059-64.
- Jaretski A 3rd, Barohn RJ, Ernstoff RM, Kaminski HJ, Keesey JC, Penn AS, et al. Myasthenia gravis: recommendations for clinical research standards. Task Force of the Medical Scientific Advisory Board of the Myasthenia Gravis Foundation of America. *Neurology* 2000; 55(1): 16-23.
- Ostovan VR, Fatehi F, Davoudi F, Nafissi S. Validation of the 15-item myasthenia gravis quality of life questionnaire (MG-QOL15) Persian version. *Muscle Nerve* 2016; 54(1): 65-70.
- Muscle Study Group. A trial of mycophenolate mofetil with prednisone as initial immunotherapy in myasthenia gravis. *Neurology* 2008; 71(6): 394-9.
- Barth D, Nabavi Nouri M, Ng E, Nwe P, Bril V. Comparison of IVIg and PLEX in patients with myasthenia gravis. *Neurology* 2011; 76(23): 2017-23.
- Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46(12): 1417-32.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000; 25(24): 3186-91.
- Sperber AD. Translation and validation of study instruments for cross-cultural research. *Gastroenterology* 2004; 126(1 Suppl 1): S124-8.
- Barohn RJ, McIntire D, Herbelin L, Wolfe GI, Nations S, Bryan WW. Reliability testing of the quantitative myasthenia gravis score. *Ann N Y Acad Sci* 1998; 841: 769-72.
- Li M, Gao Q, Yu T. Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters. *BMC Cancer* 2023; 23(1): 799.
- Martínez Pérez JA, Pérez Martín PS. Coeficiente de correlación intraclass [Intraclass correlation coefficient]. *Semergen* 2023; 49(3): 101907.
- Almanasreh E, Moles R, Chen TF. Evaluation of methods used for estimating content validity. *Res Social Adm Pharm* 2019; 15(2): 214-21.
- Madadzadeh F, Bahariniya S. Tutorial on how to calculating content validity of scales in medical research. *Perioper Care Oper Room Manag* 2023; 31: 100315.
- Farmakidis C, Pasnoor M, Dimachkie MM, Barohn RJ. Treatment of Myasthenia Gravis. *Neurol Clin* 2018; 36(2): 311-37.