Original Article

Construct Validity of the Spanish Versions of the Memorial Symptom Assessment Scale Short Form and Condensed Form: Rasch Analysis of Responses in Oncology Outpatients

Inés Llamas-Ramos, PhD, Rocío Llamas-Ramos, PhD, José Buz, PhD, María Cortés-Rodríguez, MStat, and Ana María Martín-Nogueras, PhD

ASPRODES (I.L.-R.), Salamanca, Spain; FREMAP (R.L.-R.), Salamanca, Spain; Faculty of Education (J.B.), University of Salamanca, Salamanca, Spain; Faculty of Sciences (M.C.-R.), University of Salamanca, Salamanca, Spain; and Nursing and Physiotherapy University School (A.M.M.-N.), University of Salamanca, Salamanca, Spain

Abstract

Context. The Memorial Symptom Assessment Scale (MSAS) is a self-rating instrument for the assessment of symptom distress in cancer patients. The Spanish version of the MSAS has recently been validated. However, we lack evidence of the internal construct validity of the shorter versions (short form [MSAS-SF] and condensed form [CMSAS]). In addition, rigorous testing of these scales with modern psychometric methods is needed.

Objectives. The aim of this study was to evaluate the internal construct validity and reliability of the Spanish versions of the MSAS-SF and CMSAS in oncology outpatients using Rasch analysis.

Methods. Data from a convenience sample of oncology outpatients receiving chemotherapy ($n = 306$; mean age 60 years; 63% women) at a university hospital were analyzed. The Rasch unidimensional measurement model was used to examine response category functioning, item hierarchy, targeting, unidimensionality, reliability, and differential item functioning by age, gender, and marital status.

Results. The response category structure of the symptom distress items was improved by collapsing two categories. The scales were adequately targeted to the study patients, showed overall Rasch model fit (mean Infit MnSq ranged from 0.98 to 1.05), met criteria for unidimensionality, and the reliability of scores was good (person reliability > 0.80), except for the CMSAS prevalence scale. Only four items showed differential item functioning.

Conclusion. The present study demonstrated that the Spanish versions of the MSAS-SF and CMSAS have adequate psychometric properties to evaluate symptom distress in oncology outpatients. Additional studies of the CMSAS are recommended. J Pain Symptom Manage 2018;55:1480−1491. © *2018 American Academy of Hospice and Palliative Medicine. Published by Elsevier Inc. All rights reserved.*

Key Words

Symptom distress, construct validity, oncology outpatients, Rasch model, measurement invariance

Introduction

The understanding and control of symptoms experienced by cancer patients is an issue of great importance from a clinical and research perspective. Measuring symptom distress is a challenge due to its subjective nature and because its results directly impact the patient care.

The Memorial Symptom Assessment Scale (MSAS)[1] is a widely used patient-rated instrument for

Address correspondence to: Inés Llamas-Ramos, PhD, Nursing and Physiotherapy School, University of Salamanca, C/Donantes de Sangre s/n, 37007 Salamanca, Spain. E-mails: inesllamas@usal.es or ines_llamas@hotmail.com

Accepted for publication: February 20, 2018.

0885-3924/$ - see front matter
https://doi.org/10.1016/j.jpainsymman.2018.02.017

measuring physical and psychological symptom distress in cancer patients. In recent years, there has been a growing interest in the validation of its shorter forms (the MSAS short form [MSAS-SF] and the condensed form [CMSAS])[2,3] in different languages.[4–9] To date, validation studies have been conducted within the classical test theory framework. Results have consistently found adequate psychometric properties. For example, the Spanish version of the MSAS[10] showed good validity and reliability (Cronbach's alpha = 0.80–0.90). Some authors[6] have supported the essential unidimensionality of the MSAS-SF using the nominal response model within the Item Response Theory (IRT) framework. Their findings suggested that the original response category structure should be revised.

To our knowledge, no validation study of the MSAS-SF and CMSAS has been conducted using the Rasch model. Rasch analysis has some advantages over more classical approaches[11–13] such as 1) interval properties—raw scores are transformed into linear interval-level measures of the latent construct; 2) joint measurement—the estimates of person and item parameters are represented graphically on a common metric (log odds unit or logit); 3) specific objectivity—this unique property of Rasch models means that the estimation of item parameters is independent of the persons used (i.e., person invariance) and that the estimation of the person parameters is independent of the particular items used (i.e., item invariance); 4) local independence—items are required to be independent of each other; 5) specificity of the standard error of measurement (SEM)—provides the test information function (TIF), which yields an index of the precision of the measurement at each point of the latent construct. As a result, in a growing number of studies, measuring instruments have shown significant improvement in efficiency and precision when using the IRT.[14]

Rasch analysis is adequate for guaranteeing the efficiency in measurement in health care settings where clinicians and researchers frequently have to collect health information with ever shorter scales.[15] Moreover, the Rasch model allows easy comparisons of measures and easy health status interpretation in clinical and epidemiological studies.[15,16]

The aim of this study was to evaluate the internal construct validity and reliability of the Spanish versions of the MSAS-SF and CMSAS in oncology outpatients. Specific aims of the study were to examine 1) response category functioning, 2) person-response validity, 3) item hierarchy and targeting, 4) model fit, 5) unidimensionality, 6) reliability, and 7) differential item functioning (DIF) by age, gender, and marital status.

Methods

Patients and Settings

This cross-sectional analysis is part of a previous validation study of the Spanish version of the MSAS. Details of the methods and results from the study are published elsewhere.[10] In brief, oncology outpatients were recruited using convenience sampling from the university hospital of Salamanca. Eligible patients were older than 18 years, diagnosed with any type of cancer and receiving chemotherapy treatment. Exclusion criteria were as follows: cognitive impairment, receiving the first round of chemotherapy, and difficulty in reading or writing. At the time of data gathering, patients had more cycles scheduled. For this study, we included data not available in the previous validation study with 246 patients. Sociodemographic characteristics were collected from the clinical history. A total of 637 patients were contacted, 152 did not reach the inclusion criteria, 88 met exclusion criteria, and 91 patients declined to participate. Thus, the total sample for this study comprised 306 oncology outpatients. Stable item calibrations and person measures within ±1 logit with a 99% of confidence can be obtained for polytomous items with this sample size.[17]

All participants gave written informed consent. The study was approved by the clinical research ethics committee at the university hospital of Salamanca.

Instruments

The MSAS-SF[2] measures 32 symptoms in terms of the experienced distress by the patient (28 symptoms) and the frequency of occurrence (four items). Every symptom is recorded as present (scored 1) or absent (scored 0). MSAS-SF subscales include physical symptom distress score (PHYS, 12 items), psychological symptom distress score (PSYCH, six items), the Global Distress Index (10 items), and a total distress symptom score (MSAS-SF TOTAL). Distress is rated on a five-point Likert scale (0 = not at all, 1 = a little bit, 2 = somewhat, 3 = quite a bit, 4 = very much). Frequency of occurrence is rated on a four-point Likert scale (1 = rarely, 2 = occasionally, 3 = frequently, 4 = almost constantly).

The CMSAS[3] includes only 11 items of symptom distress and three items of frequency of occurrence. Response categories are the same as for the MSAS-SF. The CMSAS yields three subscales: physical symptom distress score (CMSAS PHYS), psychological symptom frequency score (CMSAS PSYCH), and a total symptom distress score (CMSAS SUM). The prevalence of symptoms is obtained on both the MSAS-SF and CMSAS by counting the number of symptoms endorsed by the patient in a dichotomous scale.

Data Analysis

Rasch analysis is described in detail elsewhere.[13,18] Briefly, in the Rasch model, the probability of a given patient endorsing an item is a logistic function of the relative distance between the person's ability (θ) (i.e., level of symptom distress) and the item difficulty (δ) (i.e., the level of distress expressed by the item) on a conjoint linear scale. According to the model, there are two expectations[19]: 1) a patient with higher symptom distress should always have a better chance of endorsing any symptom than a healthy one and 2) any patient should always have a better chance of endorsing a frequent symptom than a more rare symptom.

Rasch analysis is based on a probabilistic model that uses joint maximum likelihood estimation to order items and persons simultaneously, thereby arranging the items along a difficulty continuum and persons along an ability continuum.

The prevalence of symptoms was examined using the Rasch model for dichotomous items. The Rating Scale Model[18,20] that is appropriate for analyzing Likert-type items was used to study the symptom distress and frequency of occurrence.

Data Preparation

The two-part question format of the MSAS-SF and CMSAS means that patients responding negatively to the first part about the presence of a symptom do not answer the second part about how distressing or frequent it was. As reported,[6] this format creates systematically missing responses that must then be replaced by other values to perform subsequent statistical analyses.

As recommended,[6,21–24] we conducted a four-stage parameter estimation approach for the two parts (distress and frequency): 1) item difficulties were estimated, 2) the fit to the Rasch model was assessed, 3) item difficulties and thresholds for the response categories were anchored (fixed at their values), and then 4) persons' abilities were estimated using the anchored item difficulties. In this way, item difficulty remains fixed to be able to compare the person level to it, while independent of their level. The steps followed for the statistical analyses and the criteria used can be seen in Table 1.

Functioning of the Response Categories

Rating scale categories should be clearly defined, be substantively relevant, and implement a conceptually exhaustive ordered sequence.[23] Reliability and validity are negatively affected by inadequate response categories.[13] Thus, we followed the recommended procedure[23] for diagnosing malfunctioning of the response categories. In addition, we investigated the person-response validity.

Item Hierarchy and Targeting

To identify the presence and level of distress associated to each symptom, the location of each symptom along the latent construct was examined. Targeting was used to examine how well the spread of items measured symptom distress or prevalence consistent with the level of patients' ability. Thus, the distribution of symptoms and patients was compared on a conjoint measure. Adequate targeting occurs when the mean person location (θ) is close to the mean item location (δ), and item locations cover a similar range of the persons as represented on the conjoint person-item map.[13,24]

Model Fit

The infit mean-square statistic (infit MnSq) that represents how accurately or predictably data fit the model[26] was used to assess the fit of each scale and the individual items. The infit statistic is sensitive to unexpected patterns of response. MnSq shows the amount of distortion of the measurement system. Its expected value is 1. Infit values over 2 indicates severe misfit.[24] If the data fit the model, then the estimation of the parameters must be invariant across subsamples of the items or persons.[19]

Unidimensionality and Local Independence

Unidimensionality was assessed conducting a principal component analysis (PCA) of the standardized residuals. PCA identifies secondary dimensions by showing unexplained relations between the item residuals after accounting for the primary Rasch dimension. A PCA on standardized residuals from a unidimensional data set is expected to extract no principal components. The results provide more insight on the question whether the physical items and the psychological items were functioning as a unidimensional construct. We also evaluated the local independence of responses by observing the correlations of the standardized residuals.

Reliability

To further determine whether the MSAS-SF and CMSAS could distinguish patients with different levels of symptom distress, person reliability (PR) and person separation (Gp)[27] were assessed. Gp represents the scale's ability to separate the patients into different levels of symptom distress. The average standard error (root mean-square error [RMSE]) was examined as part of the PR and Gp to interpret the precision of scores.[13,23,24] We also plotted the TIF and the SEM to identify how precise the MSAS-SF and CMSAS were at various ranges of the symptom distress continuum.[26]

Table 1
Analytical Steps and Statistical Criteria to Examine the Construct Validity of the Spanish Version of the MSAS-SF and CMSAS

Analytical Steps	Statistical Criteria
1. Response category functioning Examines whether patients actually discriminate between the available ordered response categories.	Guidelines indicate[24] 1) at least 10 observations from each response category 2) regular observation distribution 3) average measures advance monotonically with category 4) Outfit MnSq < 2.0 logit 5) Step calibrations advance monotonically
2. Person-response validity Provides information regarding the trustworthiness of person responses.	Person goodness-of-fit statistics[24] Infit MnSq < 2.0 logit 15% or less of misfitting persons
3. Item hierarchy and targeting Addresses the relevance and representativeness of the content of the items.	Person ability (θ) and item difficulty (δ) can be jointly represented along one dimension Values 0.6< $\lvert\theta - \delta\rvert$ <2.2 indicate good targeting for clinical samples and rating scales[13,24]
4. Model fit Indicates how accurately or predictably data fit the model.	Fit statistics indicate how accurately the data fit the model.[11,24] Infit MnSq 0.5–1.7 is adequate for clinical observations Values ≥2.0 indicate a severe misfit
5. Dimensionality and local independence Addresses the fidelity of the structure of items to the structure of the construct domain at issue.	Principal component analysis of the standardized residuals Residual variance <10% and eigenvalue <3[24,25] Inter-item correlations of the residuals <0.40 for local independence[24]
6. Reliability Indicates how well the test is able to differentiate between groups of patients with different levels of symptoms distress.	Person reliability ≥0.80 for clinical use, person separation ≥2.0, and RMSE < 0.5[13,24] TIF >4 and SEM <0.5 for precise person measures
7. Differential item functioning Ensures that scale items function uniformly across various groups within a population.	DIF contrast >0.64 and delta Mantel-Haenszeld value above |1.5| and statistically significant. Bonferroni correction was used.[24,26]

MSAS-SF = Memorial Symptom Assessment Scale Short Form; MSAS-CF = Memorial Symptom Assessment Scale Condensed Form; Infit MnSq = information-weighted mean square; RMSE = root mean-square error; TIF = test information function; SEM = standard error of measurement; DIF = differential item functioning.

Differential Item Functioning

Finally, we conducted DIF analysis by age, gender, and marital status to identify items for which patients, with the same symptom distress levels but from different groups, have different probability of endorsement. We calculated the differences in item calibrations between groups and tested the differences using the Rasch method (DIF contrast) and the Mantel-Haenszeld method, adjusting for multiple comparisons. Some studies with medical patients and community samples have found gender and age bias when measuring symptoms. It has been hypothesized that psychological (e.g., coping styles), biological, and social factors, such as having a partner, can explain gender, age, and marital status–related differences in symptom labeling, description, and reporting.[28]

We performed Rasch analysis to examine 1) the responses regarding the prevalence of symptoms, 2) the validity of the rating scoring system, 3) the subscales of the MSAS-SF and CMSAS, and 4) the MSAS-SF SUM and CMSAS SUM. The analyses were conducted using WINSTEPS, version 3.69,[29] and SPSS, version 21.0 (IBM Corp., Armonk, NY).

Results

Descriptive results of the sample are summarized in Table 2. Patients were classified into groups for DIF analysis (<60 years vs. ≥60 years; male vs. female; married vs. nonmarried) (Table 2).

The MSAS-SF and CMSAS: The Prevalence Scale

Regarding the hierarchy of symptoms, Fig. 1 reveals that the easiest symptoms to endorse were "lack of energy," "worrying," and "dry mouth," whereas the most difficult were "problems with urination" and "vomiting." All items were well targeted to the patients. Infit statistics for persons revealed that their responses fitted the model's expectations in both scales. MnSq statistic for individual items was in the appropriate range (Table 3). Overall mean item fit was excellent. No single item showed severe misfit to the model. Consequently, observed data closely follow the model expectation. PCA of the residuals supported the essential unidimensionality of both scales. The residual correlation matrix did not reveal correlations indicating local dependence. The reliability of the MSAS-SF scores showed that the scale was able to differentiate three strata of patients (Table 4). The value of the standard error (RMSE = 0.48) indicated adequate precision in the model estimates of patient measures. Reliability for the CMSAS scores was lower than desirable. Moreover, the person measures were not estimated with adequate precision (RMSE = 0.70). No evidence of DIF was found by age, gender, or marital status.

Table 2
Sample Demographic and Health Information ($N = 306$)

Characteristics	
Age (yrs), mean ± SD (range)	59.8 ± 11.7 (23−85)
Gender, *n* (%)	
Men	113 (36.9)
Women	193 (63.1)
Marital status, *n* (%)	
Married	208 (68)
Single	45 (14.7)
Divorced	20 (6.5)
Widowed	33 (10.8)
Educational level, *n* (%)	
Primary school	130 (42.5)
Secondary school	40 (13.1)
High school	52 (17)
University	84 (27.4)
Employment status, *n* (%)	
Worker	31 (10.1)
Temporary disability	75 (24.5)
Housewife	57 (18.6)
Unemployed	15 (4.9)
Retired	128 (41.8)
Cancer type, *n* (%)	
Digestive	75 (25.5)
Urinary tract	11 (3.6)
Male genital organs	7 (2.3)
Lung	36 (11.8)
Breast	88 (28.8)
Gynecologic	29 (9.5)
Hematological	32 (10.5)
Head and neck	13 (4.2)
Others	15 (4.8)
Number of symptoms, mean ± SD (range)	
MSAS-SF	11.4 ± 6.1 (0−30)

MSAS-SF = Memorial Symptom Assessment Scale Short Form.

Response Category Functioning

All the statistics for the response categories of distress were adequate. The category probability curves for the five-point Likert scale (Fig. 2) show how step calibrations were properly ordered, although they revealed that patients had difficulty discriminating between the categories "somewhat" and "quite a bit," so they were collapsed. The new response category structure improved the reliability of scores. Alternative recoding procedures were also checked, but no better results were found. Response categories of frequency of occurrence fulfilled the established criteria, and no modifications were carried out. Person fit was good, and the number of persons not responding as expected by the model was acceptable (10.4%) (Fig. 2).

Analysis of the Subscales

For the MSAS-SF, the spread of physical items (PHYS) (δ from −0.94 to 1.41 logits) was smaller than the range of patient's symptom distress (θ from −3.36 to 1.10), but the subscale was well targeted ($|\theta - \delta| = 1.85$ logits). A similar picture of item hierarchy and targeting was found for the PSYCH subscale (δ from −1.36 to 1.09 logits; θ from −3.99 to 5.23, and $|\theta - \delta| = 1.85$ logits) and the Global Distress Index (δ from −0.79 to 1.16 logits; θ from −3.48 to 2.53, and $|\theta - \delta| = 1.39$ logits). Overall, the results indicated that the subscales were better targeted to patients with moderate to high level of symptom distress.

The CMSAS subscales did not show noticeable gaps in their person-item maps, but the narrow spread of items of the PSYCH subscale (δ from −1.10 to 0.82 logits; θ from −4.40 to 6.03) indicated a problem of underrepresentation of the latent construct.

Analysis of the items of the MSAS-SF and CMSAS subscales revealed that the item "weight loss" did not demonstrate acceptable fit to the model, meaning that the patients' scores on this particular item were inconsistent with their overall response patterns. This item was not degrading the measure (Table 5). The PCA of the residuals did not identify substantial subsets of items with a strength of at least three items and explaining more than 10% of the residual variance. The examination of the item residuals further supported the local independence of the items. The reliability of scores from the subscales achieved the expected value ($PR \geq 0.80$) indicating they were able to differentiate at least three strata of patients. However, the CMSAS PSYCH was only able to differentiate two strata ($PR = 0.72$), and their person estimates were highly imprecise ($RMSE = 0.96$) (Table 4).

The MSAS-SF SUM and CMSAS SUM

The MSAS-SF SUM person-item map demonstrates that the person's measures showed great dispersion ($\theta = -3.93$ to 0.67 logits) (Fig. 3). The symptom distress measures ranged from $\delta = -0.62$ logits (i.e., most easy to endorse: "hair loss") to $\delta = 1.11$ logits (i.e., most difficult to endorse: "weight loss"), while the symptom distress for CMSAS SUM items ranged from $\delta = -0.83$ logits (i.e., most easy to endorse: "constipation") to $\delta = 1.28$ logits (i.e., most difficult to endorse: "weight loss"). The SE of the items indicated that the locations were measured with adequate precision in both scales. The mean of patients was at a lower level ($\theta = -1.63$ logits) than the mean of items indicating that overall the sample had lower symptom distress than the MSAS-SF SUM item target. Similar results were found for the CMSAS SUM (Fig. 3).

The item fit statistics indicated that only two items showed misfit: "hair loss" (Infit MnSq = 1.96) within the MSAS-SF SUM and "weight loss" (Infit MnSq = 1.93) within the CMSAS SUM. The PCAs demonstrate the essential unidimensionality of both scales. Residual correlations did not show the presence of local dependency.

The person reliability yielded very precise estimates showing that the MSAS-SF SUM can differentiate four levels of symptom distress in this sample. The graphical representation of the TIF and SEM revealed that

```
 3            .  |
                 |
                 |
              .  |
                 |
             .#  |
                 |
 2            .  |
                 |Ti Urination
                 |
              .  |
             Tp|    Vomiting
              .  |
              #  |    Dizziness     Swallowing     I don´t look like my self
 1           .#  |Si
             ##  |    Itching
                 |    Diarrhea    Swelling arms/legs     Shortness of breath
              .  |    Mouth sores
           .###  |    Weight loss       Sweats      Cough
           .###Sp|    Bloated     Nausea       Changes in skin
          .####  |    Difficulty concentrating         Drowsy
 0       .#####  |Mi Lack of appetite       Problems sexual interest/activity
          .####  |
          .####  |
        .######  |    Hair loss
           .###  |    Irritable     Sad
     .#########  |
       ######Mp|    Change food tastes      Pain    Difficulty sleeping   Constipation
-1               |Si Numbness/tingling hands/feet       Nervous
          .####  |
        .######  |    Dry mouth
         ######  |
                 |
       ########  |
             Sp|Ti  Worrying
-2       .#####  |
                 |
           .###  |
                 |    Lack of energy
           .###  |
                 |
                 |
-3           Tp|
            ###  |
                 |
                 |
                 |
                 |
              .  |
Each "#" represents 3 patients. Each "." is 1 or 2 patients
```

Fig. 1. Person-item map of prevalence of symptoms. In bold, psychological symptoms according with the authors of the scale. On the map, both patients (left-hand side) and symptoms (right-hand side) are located along a single scale. Estimates toward the top of the map represent patients with a higher number of symptoms as well as symptoms more difficult to endorse. Patients with a lower number of symptoms and the easiest symptoms to endorse are placed at the bottom. The means of patient's measures and symptom's locations are shown as the corresponding Ms on the map. The Ss (Sp for persons and Si for items) and Ts (Tp for persons and Ti for items) represent ±1 and ±2 SD, respectively. As can be seen, symptoms such as "lack of energy" and "worrying" were easy to endorse even for patients with low prevalence scores (around five symptoms, located at −1 SD). However, "itching" was difficult to endorse for them but easier to endorse for patients with high prevalence scores.

a raw score of 43 offered maximum information (a higher peak at −0.96 logits) and minimum error (SEM = 0.17) of the symptom distress measurement. The CMSAS scores were able to separate three groups of patients. The highest peak of the information function was 12, corresponding to a raw score of 18. Measurement precision for the CMSAS SUM diminished rapidly for patients with raw scores less than 3 (SEM > 0.5) (Fig. 4).

We found evidence of DIF by age in one item of the MSAS-SF SUM. At the same level of symptom distress, the item "problems with urination" was easier to agree with for patients ≥60 years ($\theta = 0.56$) than for patients <60 years ($\theta = 1.25$). Analysis by gender revealed that symptoms such as "difficulty concentrating" and "hair loss" were easier to endorse for female ($\theta = 0.01$ and $\theta = -0.56$, respectively) than for male ($\theta = 0.64$ and $\theta = -0.03$, respectively). In the CMSAS SUM, "weight loss" was easier to endorse for male ($\theta = 0.29$) than for female ($\theta = 1.05$). No DIF by age was detected, and no DIF by marital status was found on either scale.

Discussion

Our study aimed to examine the internal construct validity and reliability of the MSAS-SF and CMSAS using the strong measurement requirements of the Rasch model. Rasch analysis revealed that the Spanish

Table 3
Rasch Fit Statistics and Rasch Item Estimates for the Prevalence Scale of the Spanish Versions of the MSAS-SF and CMSAS

	Symptom	MSAS-SF			CMSAS		
		Infit	δ	SE	Infit	δ	SE
1	Difficulty concentrating	0.89	0.19	0.14	0.87	0.76	0.14
2	Pain	0.98	0.80	0.13	1.06	−0.33	0.13
3	Lack of energy	0.83	−2.50	0.16	0.89	−2.26	0.17
4	Cough	1.17	0.42	0.14	—	—	—
5	Changes in skin	1.08	0.30	0.14	—	—	—
6	Dry mouth	1.04	−1.33	0.13	1.18	−0.92	0.14
7	Nausea	0.91	0.32	0.14	—	—	—
8	Feeling drowsy	0.95	0.10	0.14	0.97	0.66	0.14
9	Numbness/tingling in hands and feet	1.04	−0.93	0.13	—	—	—
10	Difficulty sleeping	1.08	−0.80	0.13	1.13	−0.33	0.13
11	Feeling bloated	0.93	0.23	0.14	—	—	—
12	Problems with urination	0.92	1.85	0.20	—	—	—
13	Vomiting	0.98	1.41	0.18	1.06	2.08	0.18
14	Shortness of breath	0.92	0.68	0.15	0.93	1.29	0.15
15	Diarrhea	1.10	0.65	0.15	—	—	—
16	Sweats	1.00	0.40	0.14	—	—	—
17	Mouth sores	1.04	0.57	0.15	—	—	—
18	Problems with sexual interest/activity	1.22	−0.01	0.13	—	—	—
19	Itching	1.08	0.81	0.15	—	—	—
20	Lack of appetite	1.02	−0.01	0.13	1.03	0.54	0.14
21	Dizziness	0.94	1.12	0.17	—	—	—
22	Difficulty swallowing	0.92	1.20	0.17	—	—	—
23	Change in the way food tastes	1.01	−0.89	0.13	—	—	—
24	Weight loss	1.04	0.42	0.14	1.09	1.01	0.15
25	Hair loss	1.28	−0.46	0.13	—	—	—
26	Constipation	1.11	−0.79	0.13	1.24	−0.31	0.13
27	Swelling of arms or legs	0.98	0.70	0.15	—	—	—
28	"I don't look like myself"	0.94	1.20	0.17	—	—	—
29	Feeling sad	0.87	−0.57	0.13	0.81	−0.07	0.13
30	Worrying	0.81	−1.85	0.14	0.84	−1.51	0.15
31	Feeling irritable	0.90	−0.56	0.13	—	—	—
32	Feeling nervous	0.86	−1.07	0.13	0.87	−0.62	0.14
	Item mean infit MnSq (SD)	0.99 (0.11)			1.00 (0.13)		
	RMSE	0.14			0.15		
	Eigenvalue and PCA (%)	1.8 (8.5)			2.0 (4.5)		

MSAS-SF = Memorial Symptom Assessment Scale Short Form; MSAS-CF = Memorial Symptom Assessment Scale Condensed Form; δ = location; SE = standard error; Infit MnSq = information-weighted mean-square; RMSE = root mean-square error; Eigenvalue = number of items of a potential secondary dimension; PCA = principal component analysis of standardized residuals; (%) = explained variance by a potential secondary factor.

versions of the MSAS-SF and CMSAS are robust, unidimensional measures of symptom distress largely free of DIF for age, gender, and marital status.

Table 4
Reliability of the MSAS-SF and CMSAS

	Reliability	Gp	Strata	RMSE
Prevalence				
MSAS-SF	0.83	2.22	3	0.47
CMSAS	0.72[a]	1.61	2	0.70
Symptom distress				
MSAS-SF PHYS	0.80	2.00	3	0.53
MSAS-SF PSYCH	0.85	2.36	3	0.60
MSAS-SF GDI	0.84	2.28	3	0.40
MSAS-SF SUM	0.88	2.72	4	0.24
CMSAS PHYS	0.79[a]	1.95	2	0.40
CMSAS PSYCH	0.82	2.25	3	0.96
CMSAS SUM	0.83	2.27	3	0.35

MSAS-SF = Memorial Symptom Assessment Scale Short Form; MSAS-CF = Memorial Symptom Assessment Scale Condensed Form; Gp = Rasch person separation value; Strata = (4Gp + 1/3); RMSE = root mean-square error; PHYS = physical; PSYCH = psychological; GDI = Global Distress Index.
[a]Below minimum value for clinical purposes.

The Rasch-based residuals demonstrated that the physical and psychological symptoms can be treated as components of a unidimensional construct, rather than independent dimensions. Our results support findings from a previous study conducted within the IRT paradigm, indicating that summative scores can be calculated.[6] The appropriateness of a summative score is a key concern at a conceptual and practical level.[30] As an example, on a conceptual level, the contribution to the explained variance of symptoms not included in any subscales is unknown. Some of these omitted symptoms (e.g., numbness or tingling in hands and feet) show a high level of accuracy and precision and are easy for any patients with any level of symptomatology to endorse. On a practical level, all items of the two-part scale are highly affected by missing responses (ranging from 44% to 60%) due to design of the scale. Finding the best methodological approach to solve this problem is challenging. A promising strategy was successfully used,[6] using the nominal response model. We use Rasch analysis to

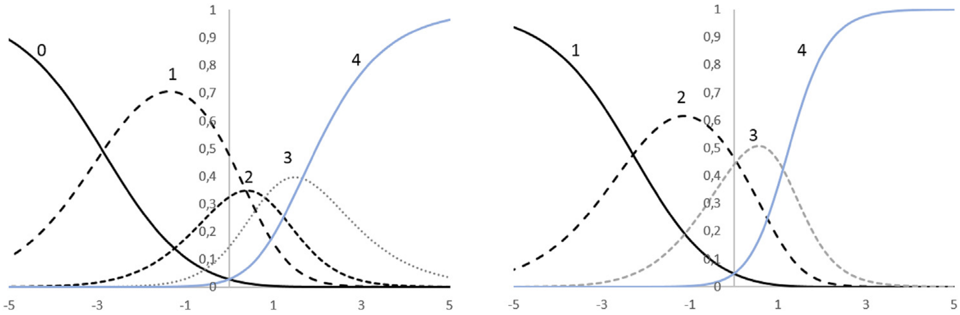

Fig. 2. Step calibrations (thresholds) are displayed as category probability curves for the 28 distress items before (left side) and after collapsing categories (right side). Curves show the probability of each category relative to the logit difference between person and item measures. Step calibrations imply that each category in turn is most likely to be chosen. Using the original rating scale (before collapsing), Categories 2 (somewhat) and 3 (quite a bit) are not distinct peaks, suggesting the difficulty of patients to choose between them. The intervals between the response Categories 2 and 3 were improved to achieve the recommended Rasch step calibrations. In addition, the response categories were recoded (1 = not at all, 2 = a little bit, 3 = quite a bit, 4 = very much).

statistically handle the problem of the missing responses and their replacements without affecting the evaluation of the quality of the scales. Thus, we provide additional evidence of the construct validity and reliability of these scales in the context of the modern psychometric methods.

Another important finding of our study is related to how the scales are adequately targeted to the population. Our results showed that the MSAS-SF and CMSAS are best targeted to patients with moderate to high levels of distress. Clinically, these patients require greater attention and, possibly, adjustments in their treatment. Therefore, these scales are fulfilling their purpose. From the patient's perspective, the symptoms "hair loss," "change in the way food tastes," and "constipation" proved to be the symptoms that cause the least distress, whereas the highest symptom distress was associated with symptoms such as "weight loss" and "feeling drowsy." We found two symptoms that do not seem to work well: "problems with sexual interest/activity" and "weight loss." Our recommendation is to keep them because they are not degrading the measure. In the case of the "weight loss" symptom, its elimination would imply a reduction of the measurement range of the latent construct. As can be seen in Figure 3, this symptom is located at the top of the item hierarchy. As for "problems with sexual interest or activity," it would be necessary to separate "interest" from "activity" because the presence of both in the same statement represents a problem for validity.

Regarding reliability, our study supports findings from previous authors and adds a new insight for future developments of the scale. Rasch analysis includes the measurement error (RMSE) associated with the person reliability. The RMSE indicated that the low reliability of the CMSAS could improve by increasing the number of items. A higher level of accuracy (e.g., RMSE < 0.50) would have indicated the need to increase the sample ability variance. However, increasing the length of the CMSAS may not be desirable for clinical settings. If the objective is to create new short versions of the original MSAS scale, it would be necessary to select the most representative symptoms and their most relevant aspects (i.e., distress, frequency, or maybe severity) and confirm the functioning of the response categories. For this purpose, the IRT approach can be very helpful.

Overall, the psychometric properties of the scales are good, but we believe there is room for improvement. First, we recommend reducing the original five-point Likert scale into a four-point Likert scale. Problems with the response categories for the symptom distress were also detected in a previous study.[6] Frequently, studies with the MSAS and its shorter forms often combine the "quite a bit" and "very much" adjacent response categories when displaying tables. According to our analysis, it is more appropriate to combine the response categories "somewhat" and "quite a bit."

Second, we recommend not deleting misfitting items. New analyses with different samples and statistical approaches would be necessary before making such an important decision. In addition, it would be interesting to examine where symptoms recently added by some authors would be located in the continuum of the construct[31] (e.g., increased appetite, weight gain, abdominal cramps, and hot flashes).

Research and Clinical Implications

The item hierarchy found by Rasch analysis in the present study is consistent with clinical expectations, is more descriptive than the summative score of symptoms, and may have more predictive value to the

Table 5
Rasch Fit Statistics and Item Estimates for the MSAS-SF and CMSAS, and Their Subscales: Physical (PHYS), Psychological (PSYCH), and the Global Distress Index (GDI)

		MSAS-SF						CMSAS		
		SUM			PHYS	PSYCH	GDI	SUM	PHYS	PSYCH
	Symptom	Infit	δ	SE	Infit	Infit	Infit	Infit	Infit	Infit
1	Difficulty concentrating	0.92	0.73	0.16	—	1.69	—	0.97	0.85	—
2	Pain	0.56	−0.38	0.12	0.69	—	0.73	0.64	0.65	—
3	Lack of energy	0.70	−0.43	0.10	0.78	—	0.89	0.83	0.79	—
4	Cough	1.30	0.45	0.17	—	—	—	—	—	—
5	Changes in skin	1.15	0.41	0.16	—	—	—	—	—	—
6	Dry mouth	0.86	−0.08	0.11	0.99	—	1.13	1.05	1.03	—
7	Nausea	0.65	0.20	0.16	0.70	—	—	—	—	—
8	Feeling drowsy	0.98	0.83	0.16	1.07	—	1.27	1.16	1.14	—
9	Numbness/tingling in hands/feet	1.02	−0.11	0.12	—	—	—	—	—	—
10	Difficulty sleeping	0.82	−0.45	0.12	—	1.19	—	0.88	0.90	—
11	Feeling bloated	0.93	0.09	0.16	1.16	—	—	—	—	—
12	Problems with urination	1.55	−0.28	0.27	—	—	—	—	—	—
13	Vomiting	0.67	0.51	0.24	0.79	—	—	0.95	0.93	—
14	Shortness of breath	0.96	0.06	0.18	—	—	—	1.13	1.13	—
15	Diarrhea	1.29	0.00	0.18	—	—	—	—	—	—
16	Sweats	0.88	0.04	0.17	—	—	—	—	—	—
17	Mouth sores	0.85	−0.10	0.17	—	—	—	—	—	—
18	Problems with sexual interest/activity	1.68	−0.53	0.15	—	—	—	—	—	—
19	Itching	1.16	0.08	0.19	—	—	—	—	—	—
20	Lack of appetite	0.93	0.25	0.15	0.93	—	1.29	1.13	1.04	—
21	Dizziness	0.98	0.05	0.21	1.08	—	—	—	—	—
22	Difficulty swallowing	0.86	−0.21	0.21	—	—	—	—	—	—
23	Change in the way food tastes	0.91	−0.56	0.12	1.02	—	—	—	—	—
24	Weight loss	1.54	1.11	0.18	1.83[a]	—	—	1.93[a]	1.83[a]	—
25	Hair loss	1.96[a]	−0.92	0.13	—	—	—	—	—	—
26	Constipation	1.12	−0.55	0.12	1.39	—	1.45	1.37	1.31	—
27	Swelling of arms or legs	1.09	−0.20	0.18	—	—	—	—	—	—
28	"I don't look like myself"	0.94	−0.63	0.21	—	—	—	—	—	—
29	Feeling sad	1.01	0.24	0.13	—	0.84	0.92	0.94	—	0.87
30	Worrying	0.98	−0.20	0.10	—	0.72	0.93	0.92	—	1.00
31	Feeling irritable	0.80	0.43	0.14	—	0.95	0.84	—	—	—
32	Feeling nervous	0.85	0.14	0.12	—	0.88	0.78	0.80	—	1.06
	Item mean Infit MnSq (SD)	1.03 (0.30)			1.03 (0.31)	1.05 (0.32)	1.02 (0.23)	1.05 (0.30)	1.05 (0.30)	0.98 (0.80)
	RMSE	0.16			0.18	0.18	0.15	0.16	0.18	0.21
	Eigenvalue and PCA (%)	2.6 (6.7)			1.9 (10.8)	1.5 (15.1)	2 (13.6)	2.1 (10.2)	1.6 (10.1)	1.6 (20.1)

MSAS-SF = Memorial Symptom Assessment Scale Short Form; MSAS-CF = Memorial Symptom Assessment Scale Condensed Form; δ = location; SE = standard error; Infit MnSq = information-weighted mean-square; RMSE = root mean-square error; Eigenvalue = number of items of a potential secondary dimension; PCA = principal component analysis of standardized residuals; (%) = explained variance by a potential secondary factor.
[a]Misfitting items (0.5 < Infit MnSq > 1.7 for unproductive measurement; infit MnSq > 2 for distorting measurement).

clinician monitoring cancer patients. For example, person-item maps of prevalence and symptom distress reveal that symptoms such as "difficulty urinating" are mainly endorsed by patients with a high number of symptoms. However, it is a symptom that causes distress among all those who suffer it.

The advantages of the Rasch model are very useful for collective assessment and individual diagnosis. Joint scale allows checking whether the scale fits the symptom distress level of the group of patients. In addition, it is possible to know if there are patient groups whose symptoms are not sufficiently represented. It also allows the clinician to better know which symptoms are representative of high levels of discomfort. At the individual level, it is possible to identify the symptoms which the person has a high or low probability of endorsing. This interpretation is richer than one based on a norm group.

The unique properties of Rasch models for item analysis allow the researcher and clinician to design shorter scales for different types of patients and in different types of settings without losing measurement precision. The Rasch model could facilitate objective comparisons of status and change at individual and group levels in cancer patients. The realization of such a goal would be of significant value to improving scientific inquiry concerning these types of patients and the therapeutical programs designed for them.

```
                            |
                            |
                            |   Weight loss
 1                          |
                            |Ti Drowsy
                        .   |   Difficulty concentrating
                        .   |   Vomiting
                            |Si Changes skin    Cough     Irritable
                        .   |   Lack appetite     Sad
                        .   |   Bloated   Nausea   I tching    Nervous
 0                          |Mi Diarrhea   Dizziness    Shortness breath     Sweats
                        .   |   Dry mouth    Numbness/tingling   Swelling   Swallowing    Mouth sores   Worrying
                      .#Tp|   Urination
                      ##    |Si Lack of energy   Pain    Difficulty sleeping
                      .#    |   Change food tastes  Constipation  Problems sexual activity  I don´t look like myself
                      ##    |
                    ####    |Ti Hair loss
-1                 #####Sp|
                 #######    |
               .########    |
                .#######    |
             .##########Mp|
             .##########    |
                  ######    |
-2             #########    |
                  .#####    |
                  .#####Sp|
                    ####    |
                      .#    |
                    .###    |
                      .#    |
-3                      Tp|
                            |
                            |
                      ##    |
                            |
Each "#" represents 3 persons.  Each "." is 1 or 2 persons
```

Fig. 3. Rasch map of patients on the 28 symptom distress items and four symptom frequency items illustrating relationship between item difficulty and person ability estimates. In bold, psychological symptoms according with the authors of the scale. Positive values indicating more symptom distress level are located above the mean for items (*Mi* on the right side of the vertical line). The Ss (Sp for persons and Si for items) and Ts (Tp for persons and Ti for items) represent ±1 and ±2 SD, respectively.

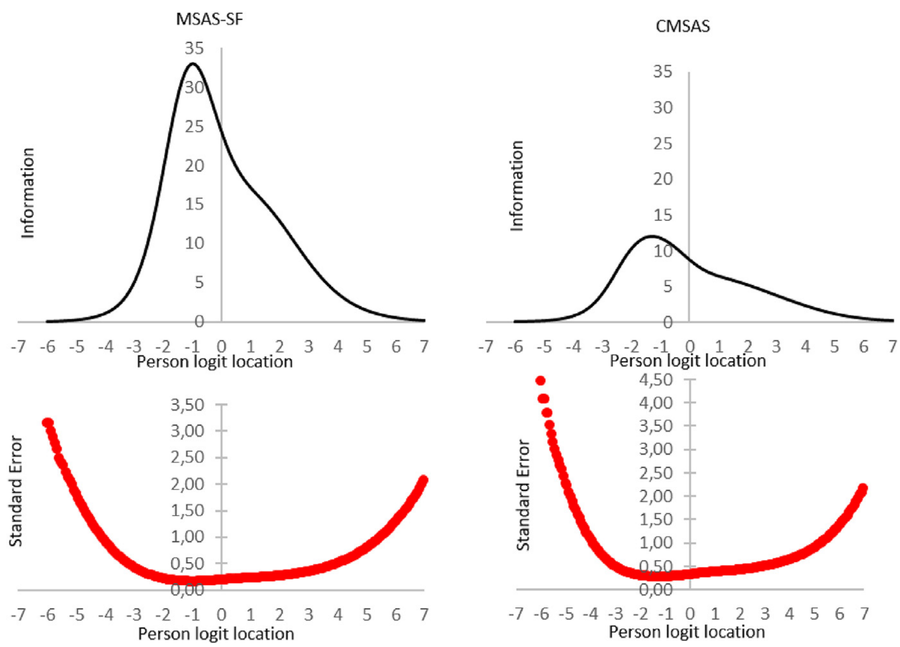

Fig. 4. At the top, the information function (y-axes) indicates the precision of the symptom distress estimates along the continuum of the latent variable. Below, the standard error of measurement (SEM; y-axes) is useful for understanding the precision of the respondent's scores. For the MSAS-SF (left side), the maximum information was 33, and the SEM was small (≤0.5), with large information (≥4), for patient scores from 3 to 123 (−3.13 to 3.87 logits) which captured 97.4% of the sample. For the CMSAS (right side), the maximum information was 12, with an optimal range for patient scores from 3 to 49 (−2.99 to 2.82 logits) which captured 92% of the sample. MSAS-SF = Memorial Symptom Assessment Scale Short Form; MSAS-CF = Memorial Symptom Assessment Scale Condensed Form.

Limitations

Our results should be interpreted considering several limitations. One is that the study is based on a convenience sample that was recruited in a single hospital. Studies with random samples from different hospitals would provide additional findings that allow a better generalization of results. Similarly, only 1% of the sample was younger than 30 years, so the generalization of our results to other age groups should be made with caution. The use of such a heterogeneous sample receiving treatment in a hospital allows us to generalize our results to real-life settings where the clinician works (ecological validity). However, we do not know if these results could be generalized to more homogeneous groups of patients with similar types of cancer. On the other hand, there may also be other patient characteristics (e.g., personality) and their family environment (e.g., social support) that we have not examined in this study and that could influence the symptom distress experience.

Disclosures and Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors have no conflicts of interest to report.

References

1. Portenoy RK, Thaler HT, Kornblith AB, et al. The Memorial Symptom Assessment Scale-an instrument for the evaluation of symptom prevalence characteristics and distress to be read. Eur J Cancer 1994;30A: 1326–1336.

2. Chang VT, Hwang SS, Feuerman M, Kasimis BS, Thaler HT. The Memorial Symptom Assessment Scale Short Form (MSAS-SF). Validity and reliability. Cancer 2000;89: 1162–1171.

3. Chang VT, Hwang SS, Kasimis B, Thaler HT. Shorter symptom assessment instruments: the Condensed Memorial Symptom Assessment Scale (CMSAS). Cancer Invest 2004; 22:526–536.

4. Lam WWT, Law CC, Fu YT, Wong KH, Chang VT, Fielding R. New insights in symptom assessment: the Chinese versions of the Memorial Symptom Assessment Scale short form (MSAS-SF) and the condensed MSAS (CMSAS). J Pain Symptom Manage 2008;36:584–595.

5. Webber K, Davies AN. Validity of the memorial symptom assessment scale-short form psychological subscales in advanced cancer patients. J Pain Symptom Manage 2011; 42:761–767.

6. Liu Y, Verkuilen J. Item response modeling of presence-severity items. Appl Psychol Meas 2013;37:58–75.

7. Dapueto JJ, Del Carmen Abreu M, Francolino C, Levin R. Psychometric assessment of the MSAS-SF and the FACIT-fatigue scale in Spanish-speaking patients with cancer in Uruguay. J Pain Symptom Manage 2014;47: 936–945.

8. Yüceege M, Sanisoğlu B, Fırat H, Ersoy Y, Sevgi E, Kurt EB. The clinical utility of Memorial Symptom Assessment-Short Form and Condensed Memorial Symptom Assessment Scale in Turkish lung cancer patients. Clin Respir J 2015;9:221–227.

9. Nho JH, Kim SR, Chang VT, Nam J-H. Reliability and validity of the Korean Memorial Symptom Assessment Scale-Short Form in gynecological cancer patients. J Pain Symptom Manage 2018;55:101–107.

10. Llamas-Ramos I, Llamas-Ramos R, Martín-Nogueras AM, et al. Reliability and validity of the Spanish version of the Memorial Symptom Assessment Scale in oncology patients. J Pain Symptom Manage 2016;52:864–891.

11. Embretson SE, Reise SP. Item response theory for psychologists. Mahwah, NJ: LEA, 2000.

12. Prieto G, Delgado AR. Análisis de un test mediante el modelo de Rasch. Psicothema 2003;15:94–100.

13. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences (3rd). Routledge. NJ: Lawrence Erlbaum Associates, Inc, 2015.

14. Las Hayas C, Bilbao A, Quintana JM, Garcia S, Lafuente I. A comparison of standard scoring versus Rasch scoring of the Visual Function Index-14 in patients with cataracts. Invest Ophthalmol Vis Sci 2011;52:4800–4807.

15. McHorney CA, Monahan PO. Postscript: applications of Rasch analysis in health care. Med Care 2004;42(1 Supp): I73–I78.

16. Buz J, Cortés-Rodríguez M. Measurement of the severity of disability in community-dwelling adults and older adults: interval-level measures for accurate comparisons in large survey data sets. BMJ Open 2016;6:e011842.

17. Wright B, Douglas G. Best test design and self-tailored testing. MESA Memorandum No. 19. Department of Education, University of Chicago, 1975.

18. Wright BD, Mok MM. An overview of the family of Rasch measurement models. In: Smith EV Jr, Smith RM, eds. Introduction to Rasch measurement. Maple Grove, MN: JAM Press, 2004:1–24.

19. Engelhard G. Invariant measurement: using Rasch models in the social, behavioral and health sciences. New York, NY: Routledge, 2013.

20. Andrich D. Rasch models for measurement. London: Sage Publications, 1988.

21. McDonald RP. Linear vs. non linear models in Item Response Theory. Appl Psychol Meas 1982;6:379–396.

22. Grice JW. Computing and evaluating factor scores. Psychol Methods 2001;6:430–450.

23. Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas 2002;3:85–106.

24. Linacre JM. A user's guide to Winsteps Ministep, Rasch-Model Computer Programs 2011. Program Manual 3.710.

25. Linacre JM, Tennant A. More about critical Eigenvalue sizes (variances) in standardized-residual principal components analysis (PCA). Rasch Meas Trans 2009;23: 1228–1232.

26. Smith EV. Understanding Rasch measurement: detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3:205–230.

27. Smith EV Jr. Evidence for the reliability of measures and the validity of measure interpretation: a Rasch measurement perspective. J Appl Meas 2001;2:281–311.

28. Barsky AJ, Peekna HM, Borus JF. Somatic symptom reporting in women and men. A review. J Gen Intern Med 2001;16:266–275.

29. Linacre JM. Winsteps-Rasch Model computer program. Version 3.70.0, 2010. Available from http://www.winsteps.com. Accessed October 27, 2017.

30. Prieto L, Alonso J, Lamarca R. Classical test theory versus Rasch analysis for quality of life questionnaire reduction. Health Qual Life Outcomes 2003;1:27.

31. Wong ML, Cooper BA, Paul SM, et al. Differences in symptom clusters identified using ratings of symptom occurrence versus severity in lung cancer patients receiving chemotherapy. J Pain Symptom Manage 2017;54:194–203.