

# Refinement and Revalidation of the Demoralization Scale: The DS-II—Internal Validity

Sophie Robinson, BPsych(Hons)<sup>1,2</sup>; David W. Kissane, MBBS, MPM, MD, FRANZCP, FACHPM, FAPM<sup>1,3,4,5</sup>;

Joanne Brooker, BSc(Hons), GradDipPsych, GradDipComp, PhD<sup>1,3</sup>;

Natasha Michael, MBChB, FRACP, FACHPM, MRCPI, MRCGP, MSc<sup>4,6</sup>;

Jane Fischer, MBBS, DipChildHealth, PGradDipMedPallCare, FACHPM<sup>7</sup>; Michael Franco, MBBS, FRACP, FACHPM<sup>5,6</sup>;

Courtney Hempton, BBS(Hons), MBioethics<sup>1,3</sup>; Merlina Sulistio, MBBS, BMedSc, FRACP, FACHPM<sup>4</sup>; Julie F. Pallant, PhD<sup>8</sup>;

David M. Clarke, MBBS, MPM, PhD, FRACGP, FRANZCP<sup>1</sup>; and Sue Burney, BA, GradDipCounsPsych, MHlthSc, PhD<sup>2,3</sup>

**BACKGROUND:** The Demoralization Scale (DS) was initially validated in 2004 to enable the measurement of demoralization in patients with advanced cancer. Subsequent shortcomings indicated the need for psychometric strengthening. Here, the authors report on the refinement and revalidation of the DS to form the DS-II, specifically reporting the scale's internal validity. **METHODS:** Patients with cancer or other progressive diseases who were receiving palliative care ( $n = 211$ ) completed a revised version of the 24-item DS and a measure of symptom burden (the Memorial Symptom Assessment Scale). Exploratory factor analysis and Rasch modeling were used to evaluate, modify, and revalidate the scale, providing information about dimensionality, suitability of response format, item fit, item bias, and item difficulty. Test-retest reliability was examined for 58 symptomatically stable patients at a 5-day follow-up. **RESULTS:** Exploratory factor analysis supported a 22-item, 2-component model. Separate Rasch modeling of each component resulted in collapsing the response option categories and removing 3 items from each component. Both final 8-item subscales met Rasch model expectations and were appropriate to sum as a 16-item total score. The DS-II demonstrated internal consistency and test-retest reliability (Meaning and Purpose subscale:  $\alpha = .84$ ; intraclass correlation [ICC] = 0.68; Distress and Coping Ability subscale:  $\alpha = .82$ ; ICC = 0.82; total DS:  $\alpha = .89$ ; ICC = 0.80). **CONCLUSIONS:** The DS-II is a 3-point response, self-report scale comprising 16 items and 2 subscales. Given its revalidation, psychometric strengthening, and simplification, the DS-II is an improved and more practical measure of demoralization for research and clinical use. External validation of the DS-II will be reported subsequently. [See editorial on pages 000-000 and companion article on pages 000-000, this issue.] *Cancer* 2016;000:000-000. © 2016 American Cancer Society.

**KEYWORDS:** psychometrics, cancer, reliability, validity, adjustment, coping behavior, demoralization, Rasch modeling.

## INTRODUCTION

Demoralization has become increasingly recognized in palliative care as a clinical issue requiring assessment and treatment.<sup>1,2</sup> Understood as a state of maladaptive coping, demoralization develops with symptoms of hopelessness and helplessness associated with loss of purpose and meaning in life.<sup>1</sup> In a recent systematic review of 25 studies, clinical prevalence rates for demoralization ranged from 13% to 18% in patients with progressive diseases like cancer.<sup>3</sup> The morale of any patient fluctuates dimensionally from optimism to mild disheartenment, to greater despondency, and potentially to deep despair, which can be associated with a desire for hastened death.<sup>4</sup> Thus, the importance of measuring demoralization has been emphasized with reference to the risk of suicide and its potential relevance in end-of-life decision making.<sup>1</sup>

Access to a psychometrically sound measure aids in the clinical assessment of demoralization.<sup>5</sup> Our preliminary validation of the Demoralization Scale (DS) in 2004 created a 24-item self-report scale that proved to be a useful measure of a poor coping response.<sup>3,5</sup> The DS was translated into several languages and was further validated with traditional classical test theory (CTT) approaches in 4 studies.<sup>2,6-8</sup> Psychometric evaluation consistently demonstrated convergent validity of the DS with established measures of psychological distress, quality of life, and desire for death, as well as strong internal

**Corresponding author:** David W. Kissane, MD, MPM, FRANZCP, FACHPM, FAPM, Department of Psychiatry, School of Clinical Sciences at Monash Health, Monash University, Level 3, P Block, Monash Medical Centre, 246 Clayton Road, Clayton, Victoria 3186, Australia. Fax: (011) 61-3-9594-6499; david.kissane@monash.edu

<sup>1</sup>Department of Psychiatry, School of Clinical Sciences at Monash Health, Monash University, Clayton, Australia; <sup>2</sup>School of Psychological Sciences, Monash University, Clayton, Australia; <sup>3</sup>Szalmuk Family Psycho-oncology Unit, Cabrini Institute, Malvern, Australia; <sup>4</sup>Cabrini Palliative Care Service, Cabrini Health, Prahran, Australia; <sup>5</sup>Supportive and Palliative Care Unit, Monash Health, Clayton, Australia; <sup>6</sup>Department of Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Australia; <sup>7</sup>Department of Palliative Care, Calvary Health Care Bethlehem, Caulfield, Australia; <sup>8</sup>Department of Rural Health, The University of Melbourne, Melbourne, Australia.

See editorial on pages 000-000 and companion article on pages 000-000, this issue.

**DOI:** 10.1002/cncr.30015, **Received:** October 28, 2015; **Revised:** February 2, 2016; **Accepted:** February 10, 2016, **Published online** Month 00, 2016 in Wiley Online Library (wileyonlinelibrary.com)

reliability for the total scale.<sup>2,5-8</sup> Discriminant validity in relation to depression was more difficult to establish.<sup>3</sup> The DS factor structure varied between 4 and 5 factors, and its test-retest reliability was not examined.<sup>3</sup> Overall, further validation was required given the inconclusive findings.<sup>5,8</sup>

To further examine the psychometric properties of the DS, the use of item response theory (IRT) models has been recommended.<sup>3</sup> In recent decades, these models have gained popularity in assisting scale development and refinement.<sup>3,9</sup> IRT techniques use mathematical models to examine the performance of each item and respondent in a scale, with the Rasch model most widely used.<sup>10</sup> In the Rasch model, dimensionality, category ordering, and item bias (differential item functioning)<sup>11</sup> are tested, while scale length can be reduced as information about items that overlap in difficulty level is provided.<sup>10</sup> From clinical experience, the length of the DS appeared to be burdensome for some patients. In addition, Rasch modeling has highlighted the limitations of reversed items, as these lead to confusion for respondents, thereby reducing the reliability of responses.<sup>12</sup>

Given these issues with the psychometric properties of the DS, scale evaluation, modification, and refinement were indicated with a palliative care population. Palliative care is delivered to patients who have progressive disease, with advanced cancer the predominate presentation. Progressive disease typically brings increased existential challenge and thus a greater risk of demoralization. We report the revalidation process in 2 parts. Here, using CTT and Rasch modeling to guide the development of the Demoralization Scale-II (DS-II), we describe its internal validity and reliability. Its external validity (convergent and divergent) is reported in a companion article by Robinson et al in this issue of *Cancer*.<sup>13</sup>

## MATERIALS AND METHODS

### **Study Design and Patients**

This was a multisite observational study, and ethical approval was obtained from each participating site's Human Research and Ethics Committee. Recruitment occurred from June 2013 to November 2014 in acute care metropolitan hospitals (Monash Health, Cabrini Health, and Calvary Health Care Bethlehem) in Melbourne, Australia. Patients were eligible if they had advanced cancer (stage IV, with a prognosis <1 year) or advanced progressive disease of any type (neurologic, cardiac, respiratory, etc) and were ineligible if they were cognitively impaired, unable to provide consent, or lacked sufficient English. Time

since diagnosis, recurrence, and currently being on treatment or being hospitalized were not inclusion criteria. The study population was typical of the profile of patients in palliative care programs. Treating physicians determined patients' eligibility.

## **Measures**

### **Sociodemographic and medical details**

Sociodemographic and medical details included primary diagnosis, illness duration, Karnofsky rating, inpatient status, age, sex, marital status, religion, educational achievement, and employment status.

### **The Memorial Symptom Assessment Scale**

The Memorial Symptom Assessment Scale (MSAS) was used to measure symptom burden by assessing the presence, frequency, severity, and associated distress of 32 symptoms (24 physical items and 8 psychological items) over the past week.<sup>14</sup> The scale has demonstrated satisfactory reliability ( $\alpha = .82$ ) and is well validated in patients with cancer.<sup>14</sup>

### **Revised Demoralization Scale**

In light of the research indicating that reversed items can reduce the reliability of responses,<sup>12</sup> consultation with an expert in the field of outcome evaluation in cancer treatment research (David Cella) was undertaken. After IRT review and confirmation, the 5 reversed items in the DS were converted to the same valence as the other 19 items. Before further scale modification, as detailed below, this revised DS contained 24 items that were rated on the original 5-point Likert scale, ranging from 0 (never) to 4 (all the time), with higher scores indicating higher levels of demoralization. The original validation of the DS identified 5 factors: Loss of Meaning and Purpose, Dysphoria, Disheartenment, Helplessness, and Sense of Failure.<sup>5</sup>

### **Procedure**

Demographic and medical data were obtained from patients' medical records. Consenting patients completed the questionnaires on their own or with assistance from a researcher. Those who were agreeable to follow-up repeated the measures approximately 5 days later to examine test-retest reliability. Repeat questionnaires were completed either face-to-face or by telephone follow-up, and patients were asked to respond in relation to how they had felt in the past day.

### **Statistical Analyses**

A sample size of 150 patients was needed to estimate item difficulty within  $\pm 0.5$  logits, with  $\alpha = .01$ , and  $\beta = .2$ .<sup>15</sup>

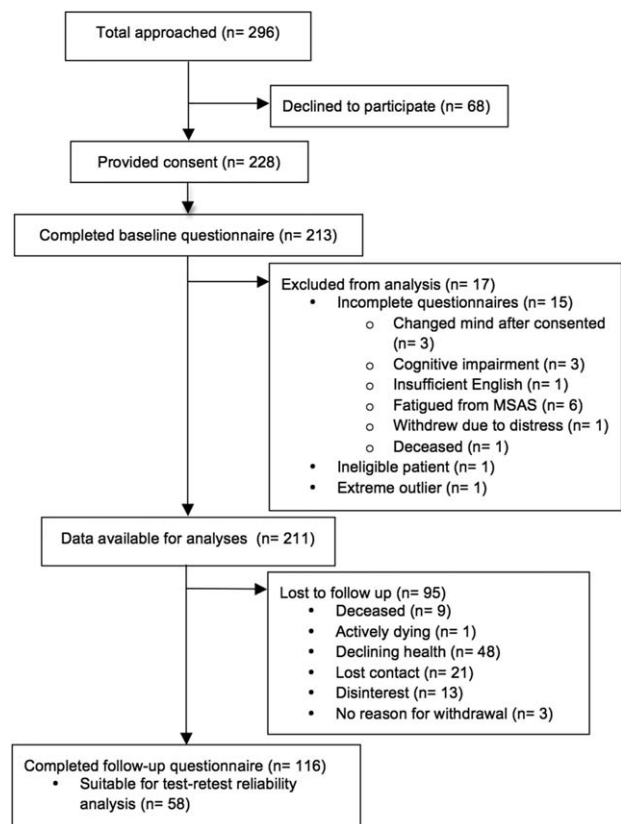
An exploratory factor analysis was undertaken using SPSS version 22 (IBM Corporation, Armonk, NY). To assure that the data were suitable for factor analysis, a Kaiser-Meyer-Olkin measure of sampling adequacy exceeding 0.6 was required<sup>16</sup> along with a significant  $\alpha$  value ( $P < .05$ ) for the Bartlett test of sphericity.<sup>17</sup> To extract the factors, a principal components analysis (PCA) was performed, with oblique rotation of the components. Three criteria were used to determine the number of factors to retain: eigenvalues  $>1$ , a Horn parallel analysis,<sup>17</sup> and inspection of the Cattell scree plot.<sup>19</sup> Software developed by Watkins (2000) was used to perform parallel analysis.<sup>20</sup>

The RUMM2030 software program (RUMM Laboratory Pty Ltd, Perth, Australia) was used to perform Rasch analyses on the subscales derived from the PCA.<sup>21</sup> Rasch analysis is a form of probabilistic testing that compares a scale against a mathematical measurement model, yielding a detailed assessment of a scale's functioning.<sup>22,23</sup> Our procedure for Rasch analysis was consistent with guidelines described elsewhere.<sup>10,24,25</sup>

To examine whether response patterns deviated from Rasch model expectations, several fit statistics were calculated.<sup>26</sup> Overall model fit was examined with a non-significant chi-square statistic, using a Bonferroni adjustment ( $P = .05/n$  items). Overall person fit and item fit were assessed by examining the summary fit residual standard deviation (SD), with a value  $<1.5$  considered acceptable.<sup>10</sup> Chi-square statistics (with a Bonferroni adjustment) and individual fit residual values were used to assess individual item fit and individual person fit. Poor fit was suggested by significant chi-square statistics or residual values  $>2.5$ . Values  $<2.5$  suggested item redundancy.<sup>26</sup>

An analysis of model misfit was also undertaken by testing for disordered thresholds, differential item functioning, and multidimensionality. We tested the suitability of the 5-point response option format by identifying disordered thresholds for items. Differential item functioning (DIF) was assessed for each item with an analysis of variance (with a Bonferroni adjustment) across sex and age. DIF is a form of item bias that occurs when groups (eg, sex, age) within the sample respond differently to an individual item despite equal levels of the underlying construct.<sup>10,27</sup>

Dimensionality was assessed by examining a PCA of the residual correlation matrix.<sup>28</sup> Subsets of items with high positive or high negative loadings were identified from the first unrotated factor in the PCA, and these subsets were compared for significant differences using a series of independent  $t$  tests.<sup>28</sup> If the lower bound of the



**Figure 1.** Patient flow is illustrated. MSAS indicates Memorial Symptom Assessment Scale.

binomial confidence interval exceeded 5% (ie, if  $>5\%$  of the tests were significant), then the scale was considered multidimensional.<sup>28</sup> Residual correlations exceeding 0.20 in the PCA of residuals were assessed for local dependency. Assessment of how well targeted the items were for participants was undertaken by examining the person-item distribution graphs.

Finally, with the objective of shortening the scale to reduce respondent burden, items appropriate for removal were identified through inspection of item maps. An item map provides information about the relative difficulty of each item.<sup>10</sup> Items listed at the same location on the map are of similar difficulty and are potential candidates for removal, when considered in conjunction with other parameters provided by the Rasch analysis and face validity.

The Rasch-derived person separation index (PSI) and the CTT-derived Cronbach  $\alpha$  statistic were used to assess internal consistency. These 2 statistics are similarly interpreted, with a value  $>.70$  considered acceptable.<sup>29</sup> Test-retest reliability was calculated with the CTT intra-class correlation (ICC) coefficient in SPSS using the 2-way random-effects design with relative agreement.<sup>30</sup>

**TABLE 1.** Sample Characteristics

Variable <sup>a</sup>	No. of Patients (%)
Total sample	211 (100)
Sex	
Men	109 (51.7)
Women	102 (48.3)
Age, Mean $\pm$ SD, y	70.98 $\pm$ 12.00
Age group, y	
40–59	44 (21)
60–79	108 (51.4)
$\geq$ 80	58 (27.6)
Marital status	
Single	24 (11.4)
Married/de facto	113 (53.5)
Divorced/separated	36 (17.1)
Widowed	38 (18)
Religion	
Christianity	116 (55.2)
Other religion	19 (9.1)
No religion	75 (35.7)
Education: Highest level achieved	
Incomplete secondary education	49 (23.4)
Secondary education	47 (22.5)
Trade or college training	51 (24.4)
Tertiary education	62 (29.7)
Employment status	
Employed	18 (8.6)
Retired	144 (68.9)
Disability pension	47 (22.5)
Recruitment site	
Cabrin Health	90 (42.6)
Calvary Health Care Bethlehem	77 (36.5)
Monash Health	44 (20.9)
Type of patient	
Inpatient	182 (86.3)
Outpatient	29 (13.7)
Primary diagnosis	
Cancer	189 (89.6)
Cardiorespiratory disease	12 (5.7)
Neurologic disease	9 (4.2)
Renal failure	1 (0.5)
Length of illness: Mean $\pm$ SD, mo	34.17 $\pm$ 45.47
Karnofsky index: Mean $\pm$ SD	56 $\pm$ 12

Abbreviations: SD, standard deviation.

<sup>a</sup>Data were missing data in some categories.

An ICC  $>0.75$  was considered *excellent*, and an ICC between 0.40 and 0.75 was deemed *fair-to-good*.<sup>31</sup> Patients who had symptoms that did not change from baseline to follow-up (indicated by a change in the MSAS score less than  $\pm 1/2$  SD)<sup>32</sup> were included in the test-retest reliability analysis.

## RESULTS

### Sample Characteristics

Of the 296 patients who were approached, 228 provided informed consent to participate (response rate, 77%). Fifteen patients were excluded because of incomplete questionnaires, 1 was excluded because of ineligibility (curative disease), and 1 was an extreme outlier aged 26 years, as illustrated in Figure 1. Of the 211 patients who

were suitable for analysis, 22 patients had a progressive disease other than cancer. The sample characteristics are summarized in Table 1.

### Scale Evaluation

#### Exploratory factor analysis

The Bartlett test of sphericity was significant ( $P < .001$ ), and the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.88, indicating suitability for factor analysis. PCA offered 5 components that had eigenvalues  $>1$ , but parallel analysis and inspection of the Cattell scree plot suggested that a 2-factor solution was optimal, explaining a total variance of 46.8%. Oblimin rotation of the 2 factors was interpretable; however, because of low loadings, we deleted 1 item from each component: item 6 (*I am not in good spirits*) and item 17 (*I am ashamed of what little I have accomplished*). The correlation ( $r$ ) between the components was 0.49. Component 1 was labeled *Meaning and Purpose* to represent the face validity of its items; similarly, component 2 was labeled *Distress and Coping Ability* to represent its items (for the component loadings, see Table 2).

#### Rasch analysis

The 2 components identified using PCA were subjected to Rasch analysis separately using the partial credit model in RUMM2030. Disordered thresholds were detected for all items on both subscales, suggesting an inappropriate response format. This was consistent with our observations during the scale administration that patients were inconsistent in their use of the options *seldom* versus *sometimes* and *often* versus *all the time*. This was resolved by collapsing response categories<sup>33</sup> to create 3 response options representing *never*, *sometimes*, and *often*. This corrected the disordered thresholds on every item for both subscales.

**Meaning and Purpose.** Rasch analysis of the 11-item Meaning and Purpose subscale had good overall fit ( $P = .01$ ; fit residual SD = 1.2), with no misfitting items or persons. No DIF was detected for sex or age. Unidimensionality of the subscale was achieved (see Table 3; analysis 1), however, local dependency was found between item 14 (*Life is no longer worth living*) and item 20 (*I would rather not be alive*; residual correlation = 0.36).

**Distress and Coping Ability.** The 11-items of the Distress and Coping Ability subscale had good fit to the Rasch model ( $P = .31$ ; fit residual SD = 0.99) with no misfitting items and 1 misfitting individual. Some degree of uniform DIF was observed by sex on item 23 (*I feel*

**TABLE 2.** Pattern Matrix of the 2-Component Principal Components Analysis Solution With Oblimin Rotation of the revised Demoralization Scale

Item	Component	
	1	2
02: My life seems to be pointless	0.83 <sup>a</sup>	-0.05
14: Life is no longer worth living	0.80 <sup>a</sup>	-0.01
03: There is no purpose to the activities in my life	0.80 <sup>a</sup>	-0.09
20: I would rather not be alive	0.74 <sup>a</sup>	-0.14
01: There is little value in what I can offer others	0.63 <sup>a</sup>	-0.06
04: My role in life has been lost	0.61 <sup>a</sup>	0.13
08: I feel that I cannot help myself	0.55 <sup>a</sup>	0.16
07: No one can help me	0.55 <sup>a</sup>	0.20
22: I feel discouraged about life	0.51 <sup>a</sup>	0.31
09: I feel hopeless	0.50 <sup>a</sup>	0.34
19: I am not a worthwhile person	0.47 <sup>a</sup>	0.15
06: I am not in good spirits	0.40	0.39
11: I feel irritable	-0.25	0.77 <sup>a</sup>
15: I tend to feel hurt easily	-0.12	0.75 <sup>a</sup>
16: I am angry about a lot of things	-0.06	0.75 <sup>a</sup>
18: I feel distressed about what is happening to me	0.08	0.68 <sup>a</sup>
24: I feel trapped by what is happening to me	0.20	0.57 <sup>a</sup>
21: I feel sad and miserable	0.26	0.57 <sup>a</sup>
10: I feel guilty	0.00	0.53 <sup>a</sup>
12: I do not cope well with life	0.27	0.51 <sup>a</sup>
13: I have a lot of regret about my life	0.14	0.48 <sup>a</sup>
05: I no longer feel emotionally in control	0.24	0.48 <sup>a</sup>
23: I feel quite isolated or alone	0.24	0.46 <sup>a</sup>
17: I am ashamed of what little I have accomplished	0.24	0.33

<sup>a</sup>These values were retained. The final version of the Demoralization Scale-II is provided in Figure 4.

*quite isolated or alone*), with a significant Bonferroni-adjusted  $\alpha$  value ( $P = .002$ ). Specifically, at equal levels of demoralization, respondents who were women endorsed a higher level on item 23 than those who were men. This minor level of DIF was consistent with clinical observations and thus was treated conservatively without further action. Unidimensionality of the subscale was supported (see Table 3, analysis 3), and there was no evidence of local dependency.

### Scale Modification

After inspection of the item maps for collocation on each subscale, items 3, 14, and 22 were removed from the Meaning and Purpose subscale; whereas items 10, 16, and 21 were eliminated from the Distress and Coping Ability subscale.

### Meaning and Purpose

The revised 8-item Meaning and Purpose subscale demonstrated improved fit to the model ( $P = .01$ ; fit residual SD = 1.02) with no misfitting items or individuals. With

the removal of item 14, local independence was observed. Nonuniform DIF by sex just reached statistical significance on item 20 (*I would rather not be alive*). Because this item has high clinical importance, a conservative approach was taken, and item 20 was retained. Unidimensionality of the scale was supported (see Table 3, analysis 2). Overall, the subscale was appropriately targeted, because there were adequate numbers of items of various difficulty to capture the distribution of respondents, as indicated in Figure 2.

### Distress and Coping Ability

Rasch analysis of the revised 8-item Distress and Coping Ability subscale indicated good overall fit ( $P = .20$ ; fit residual SD = 0.95) with no misfitting items or individuals. Uniform DIF by sex was present for item 23 (*I feel quite isolated or alone*;  $P = .002$ ). Unidimensionality of the scale was supported, and no local dependency was detected (see Table 3, analysis 4). Overall, the subscale was appropriately targeted, because there were sufficient items of varied difficulty to capture the spread of scores from respondents (see Fig. 3).

### Total scale

To test the appropriateness of summing subscale scores to provide a total score representing the underlying construct of demoralization, a substest analysis was conducted comparing individual estimates for the Meaning and Purpose subscale and the Distress and Coping Ability subscale (see Table 3, analysis 5). This test supported the unidimensionality of an underlying construct of demoralization. The final version of the DS-II is presented in Figure 4.

### Reliability

The internal consistency of the subscales and of the total 16-item scale is reported in Table 3. With reference to the MSAS scores at baseline (mean  $\pm$  SD, 0.89  $\pm$  0.44; range, 0-4), a subsample of patients who had stable symptoms ( $n = 58$ ; mean  $\pm$  SD interval days, 4.71  $\pm$  2.04) demonstrated test-retest reliability of the log of the DS-II total score (ICC = 0.80; 95% confidence interval [CI], 0.66-0.88) and subscales (log of Meaning and Purpose: ICC = 0.68; 95% CI, 0.45-0.81; log of Distress and Coping Ability: ICC = 0.82; 95% CI, 0.69-0.89). Because of the presence of outliers and some indication of nonnormality and heteroscedasticity, log transformations were performed before calculation of the ICCs.<sup>34</sup>

### Descriptive Statistics

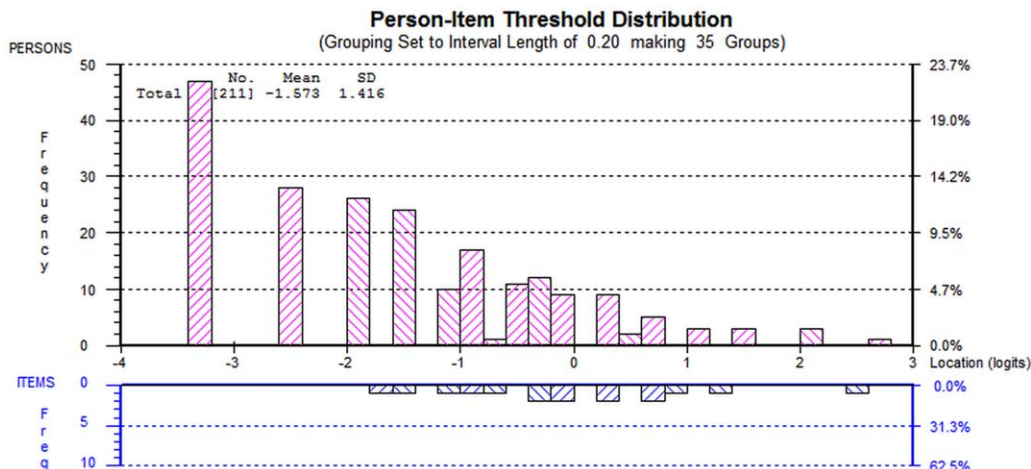
For ease of clinical utility, ordinal scores were used to report descriptive statistics. The ordinal scores in SPSS

**TABLE 3.** Summary of Results of Rasch Analyses of Demoralization Scale-II Items

Scale	Analysis	Overall Model Fit	Mean ± SD		Percentage of Significant T Tests	Internal Consistency	
			Item Fit Residual	Person Fit Residual		PSI	α
Meaning and Purpose Eleven items	1	Chi-square = 39.75; P = .01	0.13 ± 1.20	0.27 ± 1.07	3.32	.72	.89
	2	Chi-square = 31.76; P = .01	0.06 ± 1.02	0.23 ± 0.86	0.95	.64	.84
Distress and Coping Ability Eleven items	3	Chi-square = 24.78; P = .31	0.05 ± 0.99	0.28 ± 1.24	2.84	.73	.87
	4	Chi-square = 20.48; P = .20	0.04 ± 0.95	0.27 ± 1.07	1.90	.65	.82
Total: 16 items	5	Chi-square = 11.55; P = .02	-0.10 ± 0.13	-0.39 ± 0.71	0.97	.79	.89 <sup>a</sup>

Abbreviations: PSI, person separation index; SD, standard deviation.

<sup>a</sup>Cronbach α statistics for the total scale were calculated using the SPSS software package (version 22; IBM Corporation, Armonk, NY).

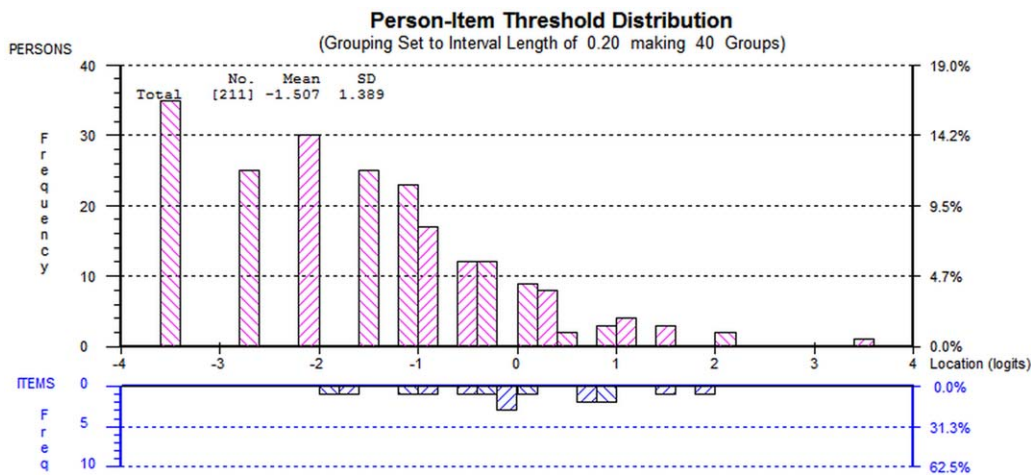


**Figure 2.** This is a person-item threshold distribution graph for the Meaning and Purpose subscale.

were converted to a 3-point scale, and items were deleted to form the 16-item scale as determined by the Rasch analyses. Summary scores for the Meaning and Purpose and Distress and Coping Ability subscales were calculated by summing the 8 items in each subscale. A total score for demoralization was calculated by summing all 16 items. Both subscales were skewed with a prominent floor effect, indicating low levels of demoralization for many in the sample tested (for a summary of the descriptive statistics, see Table 4). The Spearman correlation coefficient ( $\rho$ ) between the Meaning and Purpose and Distress and Coping Ability subscales was  $\rho = 0.61$  ( $P < .001$ ).

**DISCUSSION**

We have refined the DS to create a 16-item, 2-component scale with sound psychometric properties. Given the reduced respondent burden, the DS-II should be more user-friendly in advanced disease settings. The Meaning and Purpose subscale combines items from the Loss of Meaning and Purpose and Helplessness subscales in the original DS into a single factor.<sup>5</sup> This subscale will be a useful response measure in meaning-centered therapies. Similarly, the Distress and Coping Ability subscale combines items that previously formed the Dysphoria, Disheartenment, and Sense of Failure subscales of the original DS.<sup>5</sup> This subscale will likely be a good indicator



**Figure 3.** This is a person-item threshold distribution graph for the Distress and Coping Ability subscale.

**TABLE 4.** Descriptive Statistics for the Demoralization Scale

Variable	Meaning and Purpose	Distress and Coping Ability	Total
Mean $\pm$ SD	3.75 $\pm$ 3.67	3.89 $\pm$ 3.45	7.64 $\pm$ 6.43
Median (IQR)	3 (1–6)	3 (1–6)	6 (3–11)
Observed range	0–15	0–16	0–31
Possible range	0–16	0–16	0–32
Skewness	1.02	1.06	1.03
Kurtosis	0.32	0.77	0.70

Abbreviations: IQR, interquartile range; SD, standard deviation.

of response to cognitive and supportive therapies. The shared variance between the 2 new subscales was 36%, indicating that, although related, the 2 components measure different aspects of demoralization.

Disordered thresholds indicated that respondents were unable to reliably differentiate the 5 original response options,<sup>10</sup> yet they did so satisfactorily when 3 options (never, sometimes, or often) were used. Item reduction was possible because both components satisfied Rasch criteria for model fit, with only minor deviations. This allowed for clinical judgment to be considered in conjunction with psychometric findings. An additional 3 items from each subscale were removed, retaining model fit and resulting in an instrument consisting of 16 items with two 8-item subscales.

With regard to reliability, the DS-II demonstrated good internal consistency when measured with Cronbach  $\alpha$  statistics. The IRT-derived PSI was lower because it was affected by the skew in the data<sup>10</sup> (several respondents reported zero scores). The magnitude of Cronbach  $\alpha$  was unaffected in this manner and thus is the more relevant measure to cite. The scale demonstrated test-retest reli-

ability among a subset of patients ( $n = 58$ ) who had stable symptoms over time. When measuring test-retest reliability in a palliative care population, symptom stability is more important than precision of time interval.<sup>35</sup> Although the demonstration of test-retest reliability is an important contribution to the validation of the DS-II, these findings should be replicated in larger samples in the future.

The combination of CTT and Rasch modeling has enabled a comprehensive assessment of the DS-II. We anticipate that the DS-II will be clinically useful as an observational measure of demoralization and a measure of change in trials of meaning-based interventions. Nonetheless, there are limitations to the current findings. Although Rasch analysis is distribution free, it is affected by the spread of respondents across the construct.<sup>10</sup> Our sample was relatively small and skewed, impacting the PSI. However, this likely reflects the distribution of the construct in the targeted population. Given the sample size, we were unable to apply confirmatory factor analysis to the current data set. For clinical utility, we used the ordinal data to report descriptive statistics. In future

For each statement below, please indicate how much (or how strongly) you have felt this way **over the last two weeks** by circling the corresponding number.

		Never	Sometimes	Often
1	There is little value in what I can offer others.	0	1	2
2	My life seems to be pointless.	0	1	2
3	My role in life has been lost.	0	1	2
4	I no longer feel emotionally in control.	0	1	2
5	No one can help me.	0	1	2
6	I feel that I cannot help myself.	0	1	2
7	I feel hopeless.	0	1	2
8	I feel irritable.	0	1	2
9	I do not cope well with life.	0	1	2
10	I have a lot of regret about my life.	0	1	2
11	I tend to feel hurt easily.	0	1	2
12	I feel distressed about what is happening to me.	0	1	2
13	I am not a worthwhile person.	0	1	2
14	I would rather not be alive.	0	1	2
15	I feel quite isolated or alone.	0	1	2
16	I feel trapped by what is happening to me.	0	1	2

#### Scoring Instructions:

Total score demoralization: Sum all 16 items.

Meaning and Purpose subscale: Sum items 1, 2, 3, 5, 6, 7, 13, and 14.

Distress and Coping Ability subscale: Sum items 4, 8, 9, 10, 11, 12, 15, and 16.

**Figure 4.** Demoralization Scale-II.

research, larger studies will enable the development of Rasch conversion tables so that Rasch-derived interval data can be used and parametric tests can be utilized with confidence. Nonetheless, the difficulty of accessing large numbers of patients in a palliative care cohort is recognized.<sup>36</sup>

Overall, the DS-II is a 16-item, 2-component scale (see Figure 4) that has demonstrated item fit, unidimensionality, and reliability as a measure of demoralization in patients receiving palliative care. An examination of its external construct validity with sociodemographic factors and concurrent measures, including quality of life, depression, and attitudes toward the end of life, is reported in the companion article<sup>13</sup> and completes the revalidation process. The DS-II is likely to be a useful clinical and research tool in meaning-centered therapies and when patient populations are at risk of demoralization (eg, advanced and serious medical disease, alcohol and substance dependence, chronic mental illness, and low socioeconomic groups).

#### FUNDING SUPPORT

This work was supported by the Bethlehem Griffiths Research Foundation.

#### CONFLICT OF INTEREST DISCLOSURES

The authors made no disclosures.

#### AUTHOR CONTRIBUTIONS

**Sophie Robinson:** Conceptualization, methodology, validation, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, and project administration. **David. W. Kissane:** Conceptualization, methodology, validation, formal analysis, investigation, resources, writing—original draft, writing—review and editing, visualization, supervision, project administration, and funding acquisition. **Joanne Brooker:** Conceptualization, methodology, validation, formal analysis, writing—review and editing, and supervision. **Natasha Michael:** Conceptualization, methodology, investigation, writing—review and editing, visualization, supervision, and project administration. **Jane Fischer:** Conceptualization, methodology, resources, and writing—review and editing. **Michael Franco:** Conceptualization, methodology, investigation, writing—review and editing, and supervision. **Courtney Hempton:** Investigation, data curation, and project administration. **Merlina Sulistio:**



Conceptualization, methodology, investigation, writing—review and editing, visualization, supervision, and project administration.

**Julie F. Pallant:** Methodology, formal analysis, writing—original draft, writing—review and editing, visualization, and supervision.

**David M. Clarke:** Conceptualization, methodology, validation, formal analysis, writing—review and editing, and supervision.

**Sue Burney:** Conceptualization, writing—review and editing, and supervision.

## REFERENCES

- Kissane DW, Clarke DM, Street AF. Demoralization syndrome: a relevant psychiatric diagnosis for palliative care. *J Palliat Care*. 2001; 17:12-21.
- Mehnert A, Vehling S, Hocker A, Lehmann C, Koch U. Demoralization and depression in patients with advanced cancer: validation of the German version of the Demoralization Scale. *J Pain Symptom Manage*. 2011;42:768-776.
- Robinson S, Kissane DW, Brooker J, Burney S. A systematic review of the demoralization syndrome in individuals with progressive disease and cancer: a decade of research. *J Pain Symptom Manage*. 2015;49:595-610.
- Robinson S, Kissane DW, Brooker J, Burney S. A review of the construct of demoralization: history, definitions, and future directions for palliative care. *Am J Hosp Palliat Med*. 2016;33:93-101.
- Kissane DW, Wein S, Love A, Lee XQ, Kee PL, Clarke DM. The Demoralization Scale: a report of its development and preliminary validation. *J Palliat Care*. 2004;20:269-276.
- Costantini A, Picardi A, Brunetti S, et al. Italian version of Demoralization Scale: a validation study [in Italian]. *Riv Psichiat*. 2013;48: 234-239.
- Hung H, Chen H, Chang Y, et al. Evaluation of the reliability and validity of the Mandarin version of Demoralization Scale for cancer patients. *J Intern Med Taiwan*. 2010;21:427-435.
- Mullane M, Dooley B, Tiernan E, Bates U. Validation of the Demoralization Scale in an Irish advanced cancer sample. *Palliat Support Care*. 2009;7:323-330.
- Luo X, Cappelleri JC, Cella D, et al. Using the Rasch model to validate and enhance the interpretation of the Functional Assessment of Cancer Therapy—Kidney Symptom Index—Disease-Related Symptoms scale. *Value Health*. 2009;12:580-586.
- Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol*. 2007;46:1-18.
- Holland PW, Wainer H. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates; 2012.
- Conrad KJ, Wright BD, McKnight P, McFall M, Fontana A, Rosenheck R. Comparing traditional and Rasch analyses of the Mississippi PTSD scale: revealing limitations of reverse-scored items. *J Appl Meas*. 2004;5:15-30.
- Robinson S, Kissane DW, Brooker J, et al. Refinement and revalidation of the Demoralization Scale: The DS-II—external validity. *Cancer*. 2016;000:000-000.
- Portenoy RK, Thaler HT, Kornblith AB, et al. The Memorial Symptom Assessment Scale: an instrument for the evaluation of symptom prevalence, characteristics and distress. *Eur J Cancer*. 1994; 30:1326-1336.
- Linacre JM. Sample size and item calibration (or person measure) stability [serial online]. *Rasch Meas Transact*. 1994;7:328. Available at: <http://www.rasch.org/rmt/rmt74m.htm>. Accessed February 2015.
- Kaiser HF. A second generation little jiffy. *Psychometrika*. 1970;35: 401-415.
- Bartlett M. A note on multiplying factors for various chi-squared approximations. *J R Stat Soc B*. 1954;16:296-298.
- Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika*. 1965;30:179-185.
- Cattell RB. The scree test for the number of factors. *Multivariate Behav Res*. 1966;1:245-276.
- Watkins MW. Monte Carlo PCA for Parallel Analysis [computer software]. State College, PA: Ed & Psych Associates; 2000.
- Andrich D, Sheridan B, Luo G. Rasch Models for Measurement: RUMM2030 [computer software]. Perth, Western Australia: RUMM Laboratory Pty Ltd; 2010.
- Pallant JF, Haines HM, Hildingsson I, Cross M, Rubertsson C. Psychometric evaluation and refinement of the Prenatal Attachment Inventory. *J Reprod Infant Psychol*. 2014;32:112-125.
- Perline R, Wright BD, Wainer H. The Rasch model as additive conjoint measurement. *Appl Psychol Meas*. 1979;3:237-255.
- Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should I look for in a Rasch paper? *Arthritis Rheum*. 2007;57: 1358-1362.
- Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. *Int J Nurs Stud*. 2009;46:380-393.
- Shea TL, Tennant A, Pallant JF. Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS) [serial online] *BMC Psychiatry*. 2009;9:21.
- Lange R, Thalbourne MA, Houran J, Lester D. Depressive response sets due to gender and culture-based differential item functioning. *Pers Individual Differences*. 2002;33:937-954.
- Smith EV JR. 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: 205-231.
- Santos JRA. Cronbach's alpha: a tool for assessing the reliability of scales [serial online]. *J Extension*. 1999;37:1-5. Available at: <http://www.joe.org/joe/1999april/tt3.php?ref>. Accessed July 2015.
- Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19:231-240.
- Fleiss JL. *Design and Analysis of Clinical Experiments*. New York: John Wiley & Sons, Inc; 1999.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407-415.
- Andrich D, De Jong JH, Sheridan B. Diagnostic opportunities with the Rasch model for ordered response categories. In: Rost J, Langehein R, eds. *Applications of Latent Trait and Latent Class Models in the Social Sciences*. Munster, Germany: Waxmann; 1997:59-70.
- Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 5th ed. Boston, MA: Pearson Education; 2007.
- Paiva CE, Barroso EM, Carnesecca EC, et al. A critical analysis of test-retest reliability in instrument validation studies of cancer patients under palliative care: a systematic review [serial online]. *BMC Med Res Methodol*. 2014;14:8.
- McWhinney IR, Bass MJ, Donner A. Evaluation of a palliative care service: problems and pitfalls. *BMJ*. 1994;309:1340-1342.