ORIGINAL ARTICLE

Rasch analysis of the Edmonton Symptom Assessment System and research implications

O. Cheifetz BScPT MSc PhD,*†
T.L. Packham OT(Reg)Ont MSc,† *and*
J.C. MacDermid BScPT MSc PhD‡§

ABSTRACT

Background

Reliable and valid assessment of the disease burden across all forms of cancer is critical to the evaluation of treatment effectiveness and patient progress. The Edmonton Symptom Assessment System (ESAS) is used for routine evaluation of people attending for cancer care. In the present study, we used Rasch analysis to explore the measurement properties of the ESAS and to determine the effect of using Rasch-proposed interval-level ESAS scoring compared with traditional scoring when evaluating the effects of an exercise program for cancer survivors.

Methods

Polytomous Rasch analysis (Andrich's rating-scale model) was applied to data from 26,645 ESAS questionnaires completed at the Juravinski Cancer Centre. The fit of the ESAS to the polytomous Rasch model was investigated, including evaluations of differential item functioning for sex, age, and disease group. The research implication was investigated by comparing the results of an observational research study previously analysed using a traditional approach with the results obtained by Rasch-proposed interval-level ESAS scoring.

Results

The Rasch reliability index was 0.73, falling short of the desired 0.80–0.90 level. However, the ESAS was found to fit the Rasch model, including the criteria for uni-dimensional data. The analysis suggests that the current ESAS scoring system of 0–10 could be collapsed to a 6-point scale. Use of the Rasch-proposed interval-level scoring yielded results that were different from those calculated using summarized ordinal-level ESAS scores. Differential item functioning was not found for sex, age, or diagnosis groups.

Conclusions

The ESAS is a moderately reliable uni-dimensional measure of cancer disease burden and can provide interval-level scaling with Rasch-based scoring. Further, our study indicates that, compared with the traditional scoring metric, Rasch-based scoring could result in substantive changes to conclusions.

1. BACKGROUND

Cancer and its treatments—including surgery, chemotherapy, radiation, and hormone therapy—are associated with decreased physical function and adverse psychological and emotional effects. Physical symptoms commonly include fatigue, nausea, pain, and decreased strength and endurance[1–4]. The importance of appreciating the consequences of cancer-related side effects is magnified as the number of people living with cancer continues to increase[5]. Evaluating the impact of cancer-related (or treatment-related) side effects on patients has evolved beyond physiologic evaluations (for example, changes in hemoglobin levels) to a more holistic view that considers the constellation of symptoms that patients present (for example, pain, nausea, and anxiety)[6].

The Edmonton Symptom Assessment System (ESAS) was developed to assess a variety of symptoms often reported by patients in the palliative care setting regardless of their specific diagnosis[7]. The ESAS consists of 10 questions evaluating symptoms commonly associated with cancer, including pain, fatigue, nausea, depression, anxiety, drowsiness, appetite, well-being, shortness of breath, and an "other" condition identified as important by the patient[7]. Each item is scored on a scale ranging from 0 to 10 (maximum score of 100 if all ESAS items are

included), with higher scores indicating a higher burden of disease. The ESAS does not claim to measure health status or quality of life; rather, it focuses on bothersome symptoms that burden cancer survivors and are likely to interfere with quality of life[7]. Clinicians are able to address patient-specific symptoms identified on the ESAS and to record change over time by plotting symptom scores on a graph. Conversely, researchers have used ESAS total scores ("symptom distress score") to evaluate the effects of interventions such as exercise[8] or a combination of symptom management strategies on overall disease burden[9].

The ESAS has been extensively studied throughout the cancer care continuum in a variety of cancer diagnostic groups, and its validity and reliability have been studied in various settings and with various patient populations[10–13]. The ESAS has also been incorporated as a "standard of care assessment" by Ontario's cancer centres, which suggests that it should be used routinely in practice and research [http://www.cancercare.on.ca/cms/one.aspx?portalId=1377&pageId=57699 (accessed February 7, 2013)]. Chang *et al.*[14] conducted a validation study relating the ESAS to the Memorial Symptom Assessment Scale and the Functional Assessment of Cancer Therapy. In a sample of 233 cancer survivors, those authors reported that the ESAS had an overall Cronbach alpha of 0.79 with a 1-day test–retest reliability (Spearman correlation) in the range 0.39–0.86, depending on the symptom measured[14]. However, in a 15-year retrospective review of ESAS validation studies, Nekolaichuk *et al.*[10] located 13 publications specifically evaluating various versions of the ESAS and concluded that, although this tool has been extensively adopted into clinical use, psychometric validation is limited in scope and that further studies are required to address the gap.

To date, the literature validating the ESAS or using the scale as an outcome measure has used mostly statistical methods such as regression, correlations, and traditional descriptive statistics such as means, standard deviations, and frequencies[6,7,14,15]. Those statistical methods are grounded in classical test theory[16] and are based on associated theoretic assumptions such as that estimates of reliability and validity apply only to the population tested, and that a greater number of questions or items reduces the variability directly attributable to random error. In contrast, Rasch analysis is a statistical method in the tradition of item-response theory. It therefore includes the assumptions that, the easier the item on a scale, the more likely it is that respondents will obtain a positive score on that item; that respondents with more ability are more likely to obtain a higher score on any given item; and that the measurements do not depend on the study population if the Rasch model can be applied (that is, items on the scale are locally independent, meaning that no single data point has influence on the value of another)[17]. Further, Rasch analysis is based on mathematical modelling that supports the use of ordinal scaling (such as the scaling used in the ESAS) for interval-level calculations (such as overall score summation) if the scale is found to fit the Rasch model[17]. Pallant and Tennant[18] published an introduction to the Rasch model that included a review of some of the model's mathematical formulation. Tennant and Conaghan[19] also summarized the Rasch measurement model and discussed why and when Rasch should be used. Bond and Fox[20] provide a basic introduction to Rasch analysis, and Velozo *et al.*[21] describe the use of Rasch to produce scale-free measurements of functional ability. Rasch analysis has also been used in the development and evaluation of physical abilities questionnaires[22,23].

To date, we are unaware of any published examinations of the measurement properties of the ESAS based on Rasch analysis. That gap is a critical one, given that most published studies that have used the ESAS have used parametric statistics whose assumptions would be violated if the scale is ordinal. The purposes of the present study were to use Rasch analysis to estimate the measurement properties of the ESAS, and to determine whether a Rasch-driven scoring metric provides results different from those of the current scoring metric in an evaluation of the effects of an exercise program for people with cancer.

2. METHODS

2.1 Participants

At the Juravinski Cancer Centre (Hamilton, ON), ESAS scores are collected electronically from patients attending (standard program of care) appointments. At electronic kiosks located at the entrance to the cancer centre, all patients sign in and score the ESAS based on their symptoms at the time of completion. Patients completing the ESAS might be visiting the cancer centre at any stage of their cancer journey, including for initial assessment before or after a cancer diagnosis, for chemotherapy or radiation therapy treatments, and for short- or long-term medical follow-up. The completed ESAS scores are forwarded to the medical team for review at the patient's appointment; the needs identified by the results guide the care provided. After permission was obtained from the joint Ethics Board of Hamilton Health Sciences and McMaster University, all ESAS questionnaires completed between November 1, 2010, and October 31, 2012, were retrieved for inclusion in the present retrospective study. All completed ESAS forms were included in the analysis. Each patient in the sample has only one ESAS as part of the current analysis.

2.2 Rasch Analysis

Rasch analysis was designed *a priori* and included these components: Andrich's rating-scale model

e187

("polytomous Rasch analysis") was used because ESAS scoring is polytomous. The fit of items and persons to the Rasch model were evaluated using InFit and OutFit values of 0.7 and 1.4[24]. The assumptions of local independence and uni-dimensionality, evaluation of differential item functioning (DIF), and person separation reliability were analyzed according to the guidelines provided by Linacre[24]. To evaluate the current measurement properties of the ESAS, item deletion or re-scoring was not performed. Response reordering was not performed in the current study[25]; however, item misfit was evaluated[24].

After the Rasch analysis, ESAS scores reported by cancer survivors participating in a community-based exercise program (CanWell program)[8] were converted to standardized scores proposed by the Rasch analysis results and reanalyzed.

2.3 Statistical Analysis

All ESAS scores were exported from electronic patient charts to an Excel database and were subsequently imported into IBM SPSS Statistics (version 20.0: IBM, Armonk, NY, U.S.A.) for analysis of demographic information, including date of birth, diagnosis type, visit date, and sex. Rasch analysis was completed using the Winsteps Rasch measurement software (version 3.80.1: Linacre JM, Beaverton, OR, U.S.A.). To explore any potential changes in the conclusion about the effects of exercise on cancer disease burden in CanWell participants, a traditional (ordinal-level scoring) repeated-measures analysis and a proposed Rasch (interval-level scoring) analysis of the ESAS scores were performed using IBM SPSS Statistics. Statistical analyses were two-sided, with significance set at $p < 0.05$.

3. RESULTS

During the study period, 26,645 completed ESAS forms were collected as already outlined. Patients identified as "non-cancer respondents" ($n = 13,178$) were those awaiting pathology confirmation of their cancer diagnosis. Of all patients completing the ESAS, 1.7% ($n = 443$) did not report their cancer diagnosis. Of those reporting a diagnosis, most were women with breast cancer (12.1%), followed by patients diagnosed with genitourinary cancers, including prostate cancer (10.2%). Table I presents full demographic information for the study sample. Regardless of the symptom measured, the most common ESAS response in the study sample was 0 (Table II).

Table III presents Rasch summary statistics. Extreme scores defined by the Rasch software were omitted by the software because they do not provide additional information about the relative difficulty of the ESAS categories, and they reduce reliability scores[24]. The mean score on the ESAS as a whole was 18.7 out of 90, which indicates that the average

TABLE I   Participant demographics

Variable	Value
ESAS score	
Mean	16±15.64
Range	0–88
Sex [n (%) women]	14,054 (52.7)
Age (years)	
Mean	64.6 (14.4)
Range	17–111
Age group [n (%)]	
<24 Years	296 (1.1)
25–34 Years	727 (2.7)
35–44 Years	1,523 (5.7)
45–54 Years	3,800 (14.3)
55–64 Years	6,195 (23.3)
65–74 Years	7,171 (26.9)
75–84 Years	5,533 (20.8)
>85 Years	1,400 (5.3)
Cancer type [n (%)]	
Breast	3,232 (12.1)
Genitourinary	2,712 (10.2)
Hematologic	1,828 (6.9)
Gastrointestinal	1,323 (5)
Gynecological	1,119 (4.2)
Skin	833 (3.1)
Lung	614 (2.3)
Head and neck	427 (1.6)
Others	334 (1.3)
Sarcoma	265 (1)
Central nervous system	213 (0.8)
Primary unknown	175 (0.7)
Non-cancer respondents[a]	13,127 (49.3)

[a]   Waiting for pathology confirmation of diagnosis.
ESAS = Edmonton Symptom Assessment System.

response on each ESAS item was roughly 2 out of 10, demonstrating that patients with cancer in this sample generally score at the low end of the response range (0–10, with 10 representing worse symptoms) on each item. A general pattern of responses at the lower end of the continuum is further supported by the negative client mean for the measure in its entirety (−0.81). Considering the separation statistic (range: 1.53–1.66) and the patient reliability (0.73) together might indicate that the ESAS is not sufficiently sensitive to distinguish between patients who score high and those who score low. Ideally, separation values would be greater than 2, and reliability would be at least 0.80—or better yet, 0.90 or greater[24]. In

TABLE II   Response frequencies for the Edmonton Symptom Assessment System

Score	Response frequency								
	Pain	*Tired*	*Nauseated*	*Depression*	*Anxious*	*Drowsy*	*Appetite*	*Wellbeing*	*Dyspnea*
Respondents (*n*)	26,537	26,587	26,558	26,582	26,520	26,528	26,554	26,489	26,579
Median	0	2	0	0	1	0	1	2	0
Mode	0	0	0	0	0	0	0	0	0
0[a]	55.9	31.9	80.8	62.3	48.1	57.1	49.4	32.3	58.4
1[a]	9.5	9.4	6.2	8.2	10.5	9.1	10.2	13.3	9.8
2[a]	7.5	10.8	3.8	7.2	9.7	7.8	8.0	11.7	7.2
3[a]	6.9	10.5	2.5	6.1	8.3	6.6	6.6	9.9	5.8
4[a]	4.6	7.2	1.6	3.6	5.0	4.1	5.1	7.1	4.0
5[a]	5.0	9.7	1.7	4.9	6.7	5.2	8.4	11.7	4.8
6[a]	3.1	5.5	0.9	2.1	3.1	2.8	3.1	3.9	2.7
7[a]	2.8	5.7	0.8	2.0	2.9	2.8	3.1	4.1	2.6
8[a]	2.3	5.1	0.6	1.7	2.7	2.3	2.6	3.0	2.2
9[a]	1.0	2.0	0.3	0.8	1.1	1.0	1.3	1.2	1.1
10[a]	0.9	1.9	0.3	0.8	1.3	0.8	1.9	1.3	1.0

[a]   Percentage of all respondents (*n*=26,645).

TABLE III   Rasch person summary statistics, excluding extreme scores (*n* = 22,871)

Statistic	Total score	Count	Measure	Model error	Infit		Outfit	
					MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
Mean	18.7	9	−0.81	0.25	0.97	0	0.98	0.1
SD	15.3			0.58	0.16	0.65	1.1	0.72	1.1
Maximum	88.0	9	2.16	0.85	5.85	5.0	9.90	6.8
Minimum	1.0	2	−2.26	0.14	0.00	−4.1	0.00	−3.9

	Real	Model
RMSE	0.32	0.30
True SD	0.48	0.49
Separation	1.53	1.66
Patient reliability	0.70	0.73

Standard error of patients mean = 0.00

MNSQ = mean squares; ZSTD = Z-standardized; SD = standard deviation; RMSE = root mean square error.

contrast, the classical test theory base calculation of the Cronbach alpha on raw patient scores yielded a reliability coefficient of 0.88. When the extreme scores were included, separation and reliability values were further reduced.

Considering item fit criteria (that is, the ability of each individual item to measure a unique level of cancer symptoms), the OutFit mean square range is expected to fall between 0.7 and 1.4[26], with values in the range 0.5–1.7 considered to be appropriate for clinical observations, and values in the range 0.5–1.5 being productive for Rasch measurement[27]. For the present study, all items but "wellbeing"

(item 8 mean square: 0.65) had an acceptable item fit within the Rasch model, indicating that we are able to predict with certain accuracy how any given client will respond to a given item. Further evidence that the items are consistently discriminative are the calculated point-measures correlations, which ranged from 0.4 to 0.74[28].

For a measure to fit the Rasch model, an important condition is that the tool in question is uni-dimensional[24] (it measures one theoretical construct). Uni-dimensionality is substantiated within the Rasch model if the unexplained variance in the first contrast is less than 4 times the total unexplained variance[24].

In the present analysis, the unexplained variance in the first contrast was 8.6% and the total unexplained variance was 45.5%. Additionally, the eigenvalue of the first construct was calculated to be 1.6 (should be less than 3 to support uni-dimensionality[24]). Those results support the contention that the ESAS measures one overall construct—that is, disease burden.

Item maps are used to indicate the degree of difficulty of each individual item relative to both the scale and the construct of interest. Figure 1 presents the ESAS item map. The figure demonstrates that the patients in our sample are less "able" than the items are "difficult" (the person mean is roughly –1, and the item mean is centred on 0). The importance of that finding is discussed later in this paper.

Another key component of Rasch analysis involves an evaluation of the actual measurement or scoring used. The ESAS score for each item is in the range 0–10, offering the respondent 11 options from which to select. Figure 2 demonstrates that patients are not utilizing or discriminating well between response options 2–7 (indicated by clustering of the responses).

The DIF was evaluated by looking at patient age, sex, and diagnostic groups (as outlined in Table I). The DIF is considered to be a factor when a DIF size of 0.43 logits or more is calculated[24]. In the present study, the sex DIF ranged from –0.16 to 0.11. All DIF logits calculated for the diagnosis groups and age groups were also less than 0.43 (data not shown).

To evaluate the research implications of developing Rasch-proposed ESAS scoring (Table IV), the ESAS scores reported by CanWell participants[8] were converted to interval-level scores and subjected to a repeated-measures analysis. Contrary to the initial results, which showed participants reporting statistically significant reductions in overall disease burden [$F_{(2, 102)}$ = 3.37; p < 0.05; power: 0.6], the reanalysis (Table V) failed to show statistically significant changes [$F_{(2, 102)}$ = 1.6, p = 0.21].

4. DISCUSSION

The present study establishes that the ESAS fits the requirements of the Rasch model in that it measures a single construct (disease burden) and that ESAS scores can be converted to an interval-level scoring metric. Our study adds to previous studies demonstrating that the ESAS has moderate measurement reliability and provides important new information about the structure and scoring of the ESAS.

Several studies have used a variety of statistical methods to investigate the reliability and validity of the ESAS[10,15]. In two studies of patients with kidney disease and of patients on hemodialysis[15,29], the reliability of the ESAS was reported to be moderate-to-high (intraclass correlation coefficient: 0.70; p < 0.01). However, in 15-year narrative review of ESAS validation studies, Nekolaichuk *et al.*[10] concluded that

```
MEASURE  Patients - MAP - Item
              <more>|<rare>
     2        .  +
                 |
                 |
                 |
              .  |
                 |
                 |
     1        .  +
              .  |
              .  |
              .  |
              .  |  NAUSEATED
              .  |T
              . T|
              .  |S
              .  |
             .#  |  DEPRESSED     SHORT-of-BREATH
     0       .#  +M DROWSY        PAIN
            .##  |  ANXIOUS       APPETITE
            .##  |
           .### S|S WELLBEING
          .####  |  TIRED
          .####  |T
          .####  |
          .####  |
          .####  |
           .### M|
    -1     .####  +
          .#####  |
            .##  |
            .##  |
            .##  |
           .### S|
                 |
           .###  |
                 |
                 |
    -2     .###  +
              T  |
                 |
                 |
            .##  |
                 |
                 |
                 |
                 |
    -3 .########### +
              <less>|<frequent>
 EACH "#" IS 78. EACH "." IS 1 TO 77
```
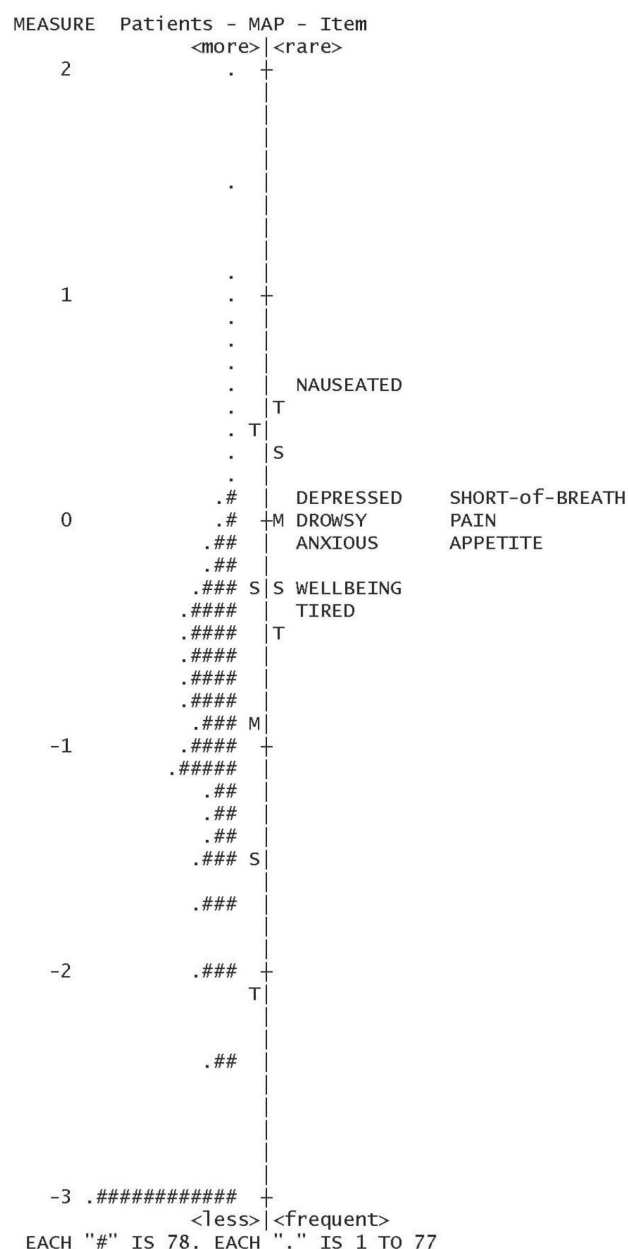
FIGURE 1 Item map of the Edmonton Symptom Assessment System (ESAS). The Rasch measure, ranging from –3 to 2 (negative values indicate either patients of lesser "ability" or items that are "easier" for the patient to endorse at high levels, and vice versa for positive values), is shown on the far left. Patients are shown immediately to the left of the vertical line, and the ESAS items, to the right. The letter "M" in the figure indicates the mean for patients and items. "S" is one standard deviation from the mean, and "T" indicates two standard deviations from the mean.

the instrument lacks the psychometric evidence for such reliability, possibly because of the various ESAS formats used in the validity studies[10], and that more validation studies are needed. Richardson and Jones[11] conducted a narrative review of the reliability and

FIGURE 2   Keyform demonstrating response consolidation around items 2–7 of the Edmonton Symptom Assessment System.

validity of the ESAS. They found 33 studies that evaluated the reliability or validity of the ESAS in patients with cancer and concluded that the ESAS is a reliable tool (correlation coefficients in the range 0.56−0.74) with restricted validity[11]. Neither review located a study evaluating the ESAS using Rasch analysis.

In the present study, the results of the Rasch analysis support use of the ESAS as a global measure of disease burden. Although the ESAS has items that might be considered to primarily evaluate physical health (pain, fatigue, nausea, drowsiness, appetite, and shortness of breath) and emotional health (depression, anxiety, well-being), our analysis demonstrated that, in patients with cancer, the ESAS total score can assess one factor: overall symptom burden. That is, the ESAS can be considered uni-dimensional. Although the symptoms measured by the ESAS are diverse, they are common complaints attributed both to the disease and to the treatment process that are often independently used to direct patient care. For example, high scores on the pain item might trigger a referral to the pain management team.

A possible question then arises: Should the ESAS be used as an evaluation of each item independently or of the total score?

The answer depends on the setting in which the ESAS is used. In the clinical setting, plotting change in the independent ESAS items over time provides the clinician with more important information than the total score does. Conversely, when evaluating the overall effect of an intervention (exercise, for example), the total ESAS score might be more relevant. Although our study has demonstrated that the ESAS is uni-dimensional, a consideration of the constructs being evaluated and of the purpose of the measurement tool is important[20,24].

We calculated the Cronbach alpha of the raw ESAS scores as 0.88, which is similar to results reported in other studies[11,14]. However, the observed patient reliability of 0.73 calculated using Rasch analysis indicates some error in how Rasch assigns person ability[24]. A possible reason for variations in ESAS reliability is that the overall scale consists of only nine items, and that study samples vary, creating some instability in the findings. Increasing the number of items might increase reliability, but might also potentially increase the response burden. Such a change would conflict with the intended purpose of the ESAS to provide a tool that is easy for patients to complete[7]. Richardson and Jones[11] suggested adding 2–3 items evaluating other symptoms to the ESAS to potentially increase the tool's reliability and to provide important additional information for the health care team. To strengthen content validity, such items could be generated from participative research. However, additions of this kind should include well-written questions that do not detract from the overall fit of the ESAS with the Rasch model, ensuring that the tool continues to be uni-dimensional in evaluating disease burden. The addition of new questions at the higher levels of the ESAS might increase the spread of the patients who respond to the ESAS, further improving its reliability.

The review by Richardson and Jones[11] discussed the fact that some patients interpreted the ESAS 11-point scale as having fewer categories. Our patient respondents had similar difficulties. As demonstrated in Figure 2, response options 2–7 are clustered for most items, indicating that they are being similarly responded to. That observation suggests that the 10 response points on the scale are too many, and that patients are not easily able to discriminate between

TABLE IV   Raw score–to–interval scale conversion, calibrated to 0–90 scoring

Score	Measure	SE	Score	Measure	SE	Score	Measure	SE
0	2.91	20.55	31	37.54	1.73	62	45.56	1.96
1	14.56	10.23	32	37.78	1.72	63	45.89	1.98
2	20.16	6.70	33	38.03	1.72	64	46.22	2.01
3	23.05	5.22	34	38.27	1.71	65	46.56	2.04
4	24.95	4.38	35	38.52	1.71	66	46.91	2.07
5	26.34	3.84	36	38.76	1.70	67	47.28	2.11
6	27.45	3.46	37	39.00	1.70	68	47.65	2.15
7	28.36	3.17	38	39.24	1.70	69	48.04	2.19
8	29.13	2.95	39	39.48	1.70	70	48.45	2.24
9	29.81	2.77	40	39.72	1.70	71	48.88	2.29
10	30.42	2.63	41	39.96	1.70	72	49.32	2.34
11	30.96	2.51	42	40.20	1.70	73	49.79	2.40
12	31.46	2.40	43	40.44	1.71	74	50.28	2.47
13	31.93	2.32	44	40.68	1.71	75	50.80	2.55
14	32.36	2.24	45	40.93	1.71	76	51.36	2.63
15	32.76	2.18	46	41.17	1.72	77	51.96	2.73
16	33.15	2.12	47	41.42	1.73	78	52.60	2.83
17	33.51	2.07	48	41.67	1.74	79	53.30	2.96
18	33.86	2.02	49	41.92	1.75	80	54.05	3.10
19	34.19	1.98	50	42.18	1.76	81	54.90	3.27
20	34.51	1.95	51	42.43	1.77	82	55.85	3.48
21	34.82	1.91	52	42.69	1.78	83	56.93	3.73
22	35.12	1.89	53	42.96	1.79	84	58.18	4.04
23	35.41	1.86	54	43.23	1.80	85	59.68	4.46
24	35.70	1.84	55	43.50	1.82	86	61.54	5.03
25	35.97	1.82	56	43.78	1.83	87	63.99	5.89
26	36.25	1.80	57	44.06	1.85	88	67.57	7.37
27	36.51	1.78	58	44.35	1.87	89	74.09	10.85
28	36.77	1.77	59	44.64	1.89	90	86.64	20.98
29	37.03	1.75	60	44.94	1.91			
30	37.29	1.74	61	45.25	1.93			

SE = standard error.

them. Future iterations of Rasch analysis of the ESAS might want to consider collapsing the response items to assess whether the fit to the model and the tool reliability change.

Rasch analysis allows for an examination of bias from individual factors such as age and sex, and an exploration of the effects on ESAS scores of a specific cancer diagnosis (for example, breast vs. prostate cancer). Although it is reasonable to assume that people with different cancers might interpret the ESAS differently, our study did not identify DIF between diagnosis groups, or sex and age groups. Given that the data for the present study were collected in an ambulatory cancer centre setting, it is unknown whether patients in a palliative setting would have a different distribution of ESAS scores. Future research might include data from the palliative setting in combination with our data to examine whether DIF is present for the two

clinical settings. Such an analysis would contribute important information to the understanding of how ESAS scores might be interpreted by different patient groups[30]. It would also support generalizability of results, because a major advantage of the Rasch tradition is that the item and person measures are not sample-dependent if the data can be shown to fit the Rasch model after adjustment for DIF[26].

The research implications of the present study are highlighted using data from the CanWell study[8]. Rasch-proposed interval-level ESAS scores (Table IV) were used to review the results of that study, in which the effects of a supervised community-based exercise program for people with cancer were evaluated. The authors reported statistically significant reductions in overall ESAS scores after completion of the 12-week exercise program[8]. However, using Rasch-produced interval-level scoring for the ESAS, a revised analysis

TABLE V Edmonton Symptom Assessment System means, current and re-scored data

Time point	Mean	SD
Before CanWell		
Current	17.9	13.68
Re-scored	−0.9354	0.71318
6–Week measure		
Current	15.4	13.36
Re-scored	−1.0542	0.73996
End CanWell		
Current	13.9[a]	11.83
Re-scored	−1.1056	0.68142

[a] Statistically significant reduction over the 12 weeks of the CanWell exercise program.

SD = standard deviation.

found no statistically significant change in ESAS scores. Although the overall conclusion of that study—that exercise does not have a negative effect on people with cancer—did not change, the conclusion that exercise reduced overall disease burden in the patient sample might have to be modified. Other researchers are encouraged to use the Rasch-proposed interval-level ESAS scores when evaluating intervention effects on the overall disease burden in cancer.

4.1 Limitations

It is important to acknowledge that, although our study demonstrated that the ESAS as a whole fits the Rasch model, some issues—such as the lower patient reliability—remain. Newer versions of the ESAS have added short explanations under items such as well-being ("wellbeing = how you feel overall") to help improve reliability and validity. Data collected in the present study antedates those changes. It is important that research and clinical settings use the newest versions of the ESAS to allow for ongoing evaluation and improvements. Future research should include Rasch analysis of data collected with the revised version of the ESAS currently used in routine care.

When the data for our study were collated, cancer stage was not available for the responding patients, limiting the ability of readers to compare their patients with ours. It is possible that patients in a palliative setting might respond differently. However, considering that one of the major advantages of Rasch analysis is that items and person measures are sample-independent within the same population, ambulatory and palliative patients would both be able to be positioned on an "ESAS ruler" and identified using DIF analysis.

5. CONCLUSIONS

Our study demonstrates that the ESAS fits the Rasch model and can be converted to an interval-level scale.

This study also supports the notion that the ESAS can be summed to produce an overall disease burden score. However, future research is needed to evaluate whether cancer survivors in different clinical settings interpret ESAS scores differently. Additionally, when interpreting ESAS scores in research, it is important to consider converting patient-reported ordinal results to interval-level scores before conducting parametric statistical testing.

6. ACKNOWLEDGMENTS

7. CONFLICT OF INTEREST DISCLOSURES

None declared.

8. REFERENCES

1. Friendenreich CM, Courneya KS. Exercise as rehabilitation for cancer patients. *Clin J Sport Med* 1996;6:237–44.
2. Courneya KS, Blanchard CM, Laing DM. Exercise adherence in breast cancer survivors training for a dragon boat race competition: a preliminary investigation. *Psychooncology* 2001;10:444–52.
3. Courneya KS, Friedenreich CM. Physical exercise and quality of life following cancer diagnosis: a literature review. *Ann Behav Med* 1999;21:171–9.
4. Lynch GS, Schertzer JD, Ryall JG. Therapeutic approaches for muscle wasting disorders. *Pharmacol Ther* 2007;113:461–87.
5. Canadian Cancer Society's Advisory Committee on Cancer Statistics. *Canadian Cancer Statistics 2013.* Toronto, ON: Canadian Cancer Society; 2013.
6. Chen E, Nguyen J, Cramarossa G, *et al.* Symptom clusters in patients with advanced cancer: sub-analysis of patients reporting exclusively non-zero ESAS scores. *Palliat Med* 2012;26:826–33.
7. Bruera E, Kuehn N, Miller MJ, Selmser P, Macmillan K. The Edmonton Symptom Assessment System (ESAS): a simple method for the assessment of palliative care patients. *J Palliat Care* 1991;7:6–9.
8. Cheifetz O, Park Dorsay J, Hladysh G, Macdermid J, Serediuk F, Woodhouse LJ. CanWell: meeting the psychosocial and exercise needs of cancer survivors by translating evidence into practice. *Psychooncology* 2014;23:204–15.

9. Modonesi C, Scarpi E, Maltoni M, *et al.* Impact of palliative care unit admission on symptom control evaluated by the Edmonton Symptom Assessment System. *J Pain Symptom Manage* 2005;30:367–73.

10. Nekolaichuk C, Watanabe S, Beaumont C. The Edmonton Symptom Assessment System: a 15-year retrospective review of validation studies (1991–2006). *Palliat Med* 2008;22:111–22.

11. Richardson LA, Jones GW. A review of the reliability and validity of the Edmonton Symptom Assessment System. *Curr Oncol* 2009;16:55.

12. Watanabe S, Nekolaichuk C, Beaumont C, Mawani A. The Edmonton symptom assessment system—what do patients think? *Support Care Cancer* 2009;17:675–83.

13. Philip J, Smith WB, Craft P, Lickiss N. Concurrent validity of the modified Edmonton Symptom Assessment System with the Rotterdam Symptom Checklist and the Brief Pain Inventory. *Support Care Cancer* 1998;6:539–41.

14. Chang VT, Hwang SS, Feuerman M. Validation of the Edmonton Symptom Assessment Scale. *Cancer* 2000;88:2164–71.

15. Davison SN, Jhangri GS, Johnson JA. Longitudinal validation of a modified Edmonton Symptom Assessment System (ESAS) in haemodialysis patients. *Nephrol Dial Transplant* 2006;21:3189–95.

16. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use.* 4th ed. New York, NY: Oxford University Press; 2008.

17. Streiner DL. Measure for measure: new developments in measurement and item response theory. *Can J Psychiatry* 2010;55:180–6.

18. 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.

19. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007;57:1358–62.

20. Bond TG, Fox CM. *Applying the Rasch Model. Fundamental Measurement in the Human Sciences.* 2nd ed. New York, NY: Routledge; 2007.

21. Velozo CA, Kielhofner G, Lai JS. The use of Rasch analysis to produce scale-free measurement of functional ability. *Am J Occup Ther* 1999;53:83–90.

22. Batcho CS, Tennant A, Thonnard JL. ACTIVLIM-Stroke: a cross-cultural Rasch-built scale of activity limitations in patients with stroke. *Stroke* 2012;43:815–23.

23. Davis AM, Perruccio AV, Canizares M, *et al.* The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. *Osteoarthritis Cartilage* 2008;16:551–9.

24. Linacre JM. *Winsteps Rasch Measurement Computer Program User's Guide.* Beaverton, OR: Winsteps.com; 2013.

25. Linacre JM. Transitional categories and usefully disordered thresholds. *Online Educ Res J* 2010;:1–10.

26. Smith E, Hudgens S. *Psychometric Evaluation of NETS Assessment Items: Phase 2.* Chicago, IL: Learning Point Associates; 2004.

27. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Meas Trans* 1994;8:370.

28. Linacre JM. The expected value of point-biserial (or similar) correlation. *Rasch Meas Trans* 2008;22:1154.

29. Davison SN, Jhangri GS, Johnson JA. Cross-sectional validity of a modified Edmonton Symptom Assessment System in dialysis patients: a simple assessment of symptom burden. *Kidney Int* 2006;69:1621–5.

30. Wright BD, Stonne MH. *Best Test Design.* Chicago, IL: MESA Press; 1979.

Correspondence to: Oren Cheifetz, Hematology/Oncology Program, Hamilton Health Sciences, Ward C4, Hamilton, Ontario L8V 1C3.
E-mail: Cheifetz@hhsc.ca

* Hematology/Oncology Program, Hamilton Health Sciences, Hamilton, ON.
† School of Rehabilitation Sciences, McMaster University, Hamilton, ON
‡ Oncology Rehabilitation Program, Hamilton Health Sciences, Hamilton, ON.
§ Hand and Upper Limb Centre, St. Joseph's Health Centre, London, ON.