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‡§

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.

KEY WORDS
Edmonton Symptom Assessment System, ESAS, Rasch analysis

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 endurance1–4. The importance of appreciating the consequences of cancer-related side effects is magnified as the number of people living with cancer continues to increase5.

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 diagnosis7. 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 patient7. Each item is scored on a scale ranging from 0 to 10 (maximum score of 100 if all ESAS items are

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.
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. 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 or a combination of symptom management strategies on overall disease burden.

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. 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. 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.

However, in a 15-year retrospective review of ESAS validation studies, Nekolaichuk et al. 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. Those statistical methods are grounded in classical test theory 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).

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. Pallant and Tennant published an introduction to the Rasch model that included a review of some of the model’s mathematical formulation. Tennant and Conaghan also summarized the Rasch measurement model and discussed why and when Rasch should be used. Bond and Fox provide a basic introduction to Rasch analysis, and Velozo et al. 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.

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

\(^a\) Waiting for pathology confirmation of diagnosis.

ESAS = Edmonton Symptom Assessment System.

Current Oncology — Volume 21, Number 2, April 2014
Copyright © 2014 Multimed Inc. Following publication in Current Oncology, the full text of each article is available immediately and archived in PubMed Central (PMC).
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\(^2\), 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\(^2\). 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\(^2\).

For a measure to fit the Rasch model, an important condition is that the tool in question is uni-dimensional\(^2\) (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\(^2\).
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\textsuperscript{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\textsuperscript{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\textsuperscript{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 [$F(2, 102) = 3.37; p < 0.05; \text{power: 0.6}$], 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\textsuperscript{10,15}. In two studies of patients with kidney disease and of patients on hemodialysis\textsuperscript{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.\textsuperscript{10} concluded that the instrument lacks the psychometric evidence for such reliability, possibly because of the various ESAS formats used in the validity studies\textsuperscript{10}, and that more validation studies are needed. Richardson and Jones\textsuperscript{11} conducted a narrative review of the reliability and...
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. 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.

We calculated the Cronbach alpha of the raw ESAS scores as 0.88, which is similar to results reported in other studies. However, the observed patient reliability of 0.73 calculated using Rasch analysis indicates some error in how Rasch assigns person ability. 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. Richardson and Jones 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 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

![Figure 2: Keyform demonstrating response consolidation around items 2–7 of the Edmonton Symptom Assessment System.](image-url)
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 if 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 if 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. 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 if.

The research implications of the present study are highlighted using data from the CanWell study. 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. However, using Rasch-produced interval-level scoring for the ESAS, a revised analysis...
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

Guidance and assistance was provided by Professor Paul Stratford and Drs. Linda Woodhouse and Julie Richardson. This study could not have been completed without statistical assistance from Christopher Condon PhD and the Winsteps Rasch measurement forum. OC was supported by the Hamilton Health Sciences Hematology Program in his role as a clinical specialist–physiotherapy and by McMaster University’s School of Rehabilitation Sciences. TLP is supported by a doctoral scholarship from the Canadian Institutes of Health Research. JCM is supported by a chair in Gender in Measurement and Rehabilitation of Musculoskeletal Work Disability from the Canadian Institutes of Health Research.

7. CONFLICT OF INTEREST DISCLOSURES

None declared.

8. REFERENCES


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.