Quantification of dyspnoea using descriptors: development and initial testing of the Dyspnoea-12

J Yorke,1 S H Moosavi,2 C Shuldham,3 P W Jones4

ABSTRACT

Rationale: Dyspnoea is a debilitating and distressing symptom that is reflected in different verbal descriptors. Evidence suggests that dyspnoea, like pain perception, consists of sensory quality and affective components. The objective of this study was to develop an instrument that measures overall dyspnoea severity using descriptors that reflect its different aspects.

Methods: 81 dyspnoea descriptors were administered to 123 patients with chronic obstructive pulmonary disease (COPD). 129 with interstitial lung disease and 106 with chronic heart failure. These were reduced to 34 items using hierarchical methods. Rasch analysis informed decisions regarding further item removal and fit to the unidimensional model. Principal component analysis (PCA) explored the underlying structure of the final item set. Validity and reliability of the new instrument were further assessed in a separate group of 53 patients with COPD.

Results: After removal of items with hierarchical methods (n = 47) and items that failed to fit the Rasch model (n = 22), 12 were retained. The “Dyspnoea-12” had good internal reliability (Cronbach’s alpha = 0.9) and fit to the Rasch model (χ² = 0.08). Items patterned into two groups called “physical” (n = 7) and “affective” (n = 5). In the separate validation study, Dyspnoea-12 correlated with the Hospital Anxiety and Depression Scale (anxiety r = 0.51; depression r = 0.44, p<0.001, respectively), 6-minute walk distance (r = -0.38, p<0.01) and MRC (Medical Research Council) grade (r = 0.48, p<0.01), and had good stability over time (intraclass correlation coefficient = 0.9, p<0.001).

Conclusion: Dyspnoea-12 fulfills modern psychometric requirements for measurement. It provides a global score of breathlessness severity that incorporates both “physical” and “affective” aspects, and can measure dyspnoea in a variety of diseases.

Dyspnoea is a perceptual experience that is complex and highly subjective. It is a common and distressing symptom in cardiopulmonary disease. Progress has been made in identifying various aspects of dyspnoea that arise from multiple factors including environmental, psychological and physiological. Other evidence indicates that dyspnoea, like pain perception, consists of “sensory quality” and “affective” components, yet no currently available dyspnoea instrument encompasses these.

Dyspnoea has been measured in two ways: (1) directly, using modified Borg or visual analogue scales, in persons experiencing dyspnoea at rest or in response to various stimuli; single item scales may not be adequate enough to capture dyspnoea complexity; and (2) indirectly, by asking the respondent to report the level of physical activity they are not able to accomplish because of dyspnoea, or within scales assessing the impact of disease on quality of life. Whilst indirect methods provide useful information, they do not quantify dyspnoea; they measure its effects on activity. This also means that they cannot measure different aspects of dyspnoea.

Patients with cardiorespiratory disease use a variety of terms to describe the experience of being breathless, and it has been proposed that dyspnoea descriptors may provide a direct route for its quantification. Studies have explored the semantics of dyspnoea, principally from a diagnostic perspective, or to understand mechanisms. That work utilised descriptor lists that focused primarily on sensory quality, but evidence suggests that sensory quality descriptors are not sufficiently robust to aid differential diagnosis, since they reflect a variety of sensations which are shared by a range of conditions. Whilst they may have limited diagnostic utility, verbal descriptors have been found to be related to the severity level of dyspnoea in patients with cardiorespiratory disease, providing a mechanism for its assessment.

The affective aspect of dyspnoea evokes distress and motivates behaviour, often overlooked in clinical and laboratory settings. This may be, in part, because no measure of this dimension currently exists. Evidence for a range of descriptors representative of affect has evolved from studies in which patients were asked to describe the experience of being breathlessness in their own words. Affect-laden descriptors were relatively consistent between studies, yet it is not known which of these may possess reliable measurement properties.

Taken together, this work has created a body of language that reflects multiple aspects of dyspnoea. A core list of dyspnoea descriptors that patients consider most relevant has not previously been defined. Our objective was to develop a concise and valid questionnaire of overall dyspnoea severity using descriptors that were relevant across different cardiorespiratory diseases. Our underlying hypothesis was that the words patients use can be applied to form a reliable scale that reflects dyspnoea severity. This paper presents two studies: the development of a new questionnaire for the quantification of dyspnoea and its initial validation.

MATERIALS AND METHODS

Development of the initial item pool

A pool of 81 items was generated from published literature reporting the language to describe breathlessness. Papers were identified by a systematic search of Medline, CINAHL and PsychInfo.
The presence of any item–trait interaction was tested using a $\chi^2$ test to assess whether all items perform consistently, regardless of overall breathlessness severity (determined by $p>0.05$). Internal reliability was assessed using Cronbach’s reliability ($\alpha$).29

**Phase 3**

Exploratory principal component analysis (PCA): PCA using varimax rotation was used to assess the underlying structure of the final item set. The number of components extracted was based on eigenvalues, and allocation of an item to a component was determined by a factor loading which by convention is set at $>0.5$.29

**Phase 4**

Preliminary validity testing of the final 12-item set: the final item set consists of 12 items. The score is calculated using simple addition of the responses for each item. It ranges from 0 to 36, where 0 represents no breathlessness and 36 represents maximal severity.

Pearson ($r$) correlations examined relationships between the 12-item set total score, MRC dyspnoea grade, Borg-intensity and Borg-distress scales. The effect of diagnosis on differences in the 12-item set score across MRC dyspnoea grades was tested using ANOVA. The average Dyspnoea-12 score and Borg-intensity and Borg-distress scores were computed for each category of MRC dyspnoea grade for the entire group ($n = 358$).

**Study two: Dyspnoea-12 validation**

The final version is called Dyspnoea-12 (Appendix I). The Dyspnoea-12 asks patients to respond to each item in relation to their breathlessness “these days” and is designed to capture the patient’s general perception about their current state, rather than a record specific to the day of testing. This time reference frame was used because it has been found to be reliable in other instruments such as the St George’s Respiratory Questionnaire.9

**Study participants and measures**

In the separate study, we tested Dyspnoea-12 reliability and validity with a further group of 53 patients with COPD. During a routine clinic visit participants completed the Dyspnoea-12, MRC dyspnoea scale; Hospital Anxiety and Depression Scale (HADS),30 lung function tests and 6-minute walk distance (6MWD). All patients repeated the Dyspnoea-12 after a median of 16 days (range 10–20). Dyspnoea-12 internal consistency was assessed using Cronbach $\alpha$, and test–retest reliability was tested using the intraclass correlation coefficient (ICC). Correlations using Pearson ($r$) were examined between Dyspnoea-12, MRC grade, HADS, forced expiratory volume in 1 s ($FEV_1$) and 6MWD.

Face validity was explored using qualitative methods (6 patients with COPD, 2 with ILD and 4 with CHF). These patients also rated the Dyspnoea-12 for ease of completion (0 = not easy to 10 = extremely easy); helpful in expressing their experience (0 = not helpful to 10 = extremely helpful); and ease of completion (0 = not easy to 10 = extremely easy).

**RESULTS**

**Study one**

A total of 358 patients completed the questionnaires; most ($n = 275$, 77%) completed these during a clinic visit, the remainder for return within 2 weeks. The majority were Caucasian ($n = 337$, 94%). Their baseline characteristics are provided in table 1.

**Phase 1**

Forty-seven items were initially removed because $\geq50\%$ of participants rated them as “none”; one item (“breathing more”) was associated with age ($p = 0.02$) (see supplementary table 1 online).
### DISCUSSION

We have developed and conducted initial validity testing of the Dyspnoea-12. It was derived from the largest pool of breathlessness descriptors that has been assembled. This is a unique instrument since it quantifies breathlessness using descriptions by patients of its qualities and its affective sequelae. We have demonstrated the concurrent validity of Dyspnoea-12 with other relevant measures. Uniquely, this instrument was developed in three disease populations. Its relationship to MRC grade was independent of diagnosis, enabling it to make direct comparisons of dyspnoea severity between patients with different diseases. Patients found the Dyspnoea-12 was easy to complete and understand, and helpful in expressing their experience of being breathless.

Unlike other self-report breathlessness instruments, Dyspnoea-12 does not depend on a reference level of activity or any specific type of activity. The reference frame “these days” is designed not to be situation specific but to be temporally specific—in terms of how patients currently experience breathlessness in their daily lives, as opposed to specifically on the day of the test or in response to a specific activity. When using this reference, patients’ responses may be determined in part by the nature and intensity of their day-to-day activities, but this possibility requires testing.

In developing this instrument, we used advanced methodology to ensure reliable measurement properties, and it represents the first application of Rasch methodology to the measurement of breathlessness. Our use of Rasch modelling facilitated the development of a questionnaire that provides measurement of dyspnoea using a parsimonious collection of items that form a unidimensional measure. Rasch analysis determined the minimum level of items required to represent the underlying construct being measured (ie, breathlessness severity). There are no rules that determine the number of items in a questionnaire; the optimum number is determined by achieving a balance between economy, precision and reliability. This is different for every instrument. Rasch methodology indicates when internal reliability begins to deteriorate with further item removal, and we found that 12 items provided the best compromise.

The 12 items contribute reliably to the measurement of overall dyspnoea severity, even though some address the affective consequences of breathlessness (eg, “makes me feel miserable”) and some measure physical aspects (eg, “more
work”). “Physical” items tended to be associated with milder severity, whereas the “affective” items were associated with more severe breathlessness. Differences in breathlessness severity associated with physical and affective items explain the patterning of items observed with PCA. These findings demonstrate the benefit of combining both physical and affective aspects in a dyspnoea scale if it is to cover a broad range of overall breathlessness severities.

We are not the first to hypothesise a multiple component dyspnoea model, since it has been suggested that a multidimensional instrument used to measure pain could be adapted to measure dyspnoea.31 Our findings are compatible with this hypothesis, since we found that a range of different descriptors from the physical to affective components of dyspnoea could be combined to form a single overall unidimensional scale. Our approach created a pragmatic, usable scale using rigorous methodology. Our primary objective was to develop an overall measure of breathlessness severity. To do this, we required all items to fit a unidimensional model. Thus, we only included “physical” and “affective” items that reflected overall severity in a similar way. This process may have excluded “physical” or “affective” items that may have behaved differently, with different measurement properties. The “physical” and “affective” components may be referred to as “components” in questionnaire parlance, but should be used only for exploratory analyses.

Patient-reported outcomes such as the Dyspnoea-12 should be derived from the words used by patients to describe disease effects and their clinical state. We are indebted to workers who collected those descriptors in previous studies. We developed our initial 81-item pool after a systematic review of existing literature and were stuck by the similarity of items identified by different research groups. Moreover, during interview sessions patients were invited to volunteer any other terms; no new terms were provided. We are confident that the Dyspnoea-12 items are relevant, appropriate and truly capture patients’ perceptions.

The terms “work” and “effort” are often used interchangeably, in part because they tended to form into one cluster.13–15 We found that the term “more work” had more reliable patient responses than “hard work” and “more effort”. Likewise, the term “panic” is often associated with the perception of breathlessness,21 22 32 yet in this large patient cohort responses to this item were erratic, so it was removed. Two of the final 12 items did demonstrate DIF associated with diagnosis. We chose to retain these because the effect of this bias on the overall score was negligible. Furthermore, the descriptors “distressing” and “difficulty catching breath” have not previously been noted as being specific to any particular disease, so this may have been a chance finding.

Figure 1 Illustration of logit (ie, severity) value for each item. *Indicates item loading onto the “physical” component identified with principal component analysis (PCA). †Indicates item loading onto the “affective” component identified with PCA. The value attached to each descriptor is the logit (ie, severity) associated with that descriptor (negative values indicate milder, positive values indicate more severe). By convention, in Rasch analysis the severity scale is centred on zero logits. Note: these logit values are not used to score the instrument; they are shown here to illustrate the different relative severity of the items.

Figure 2 Mean (+95%CI) of total Dyspnoea-12 scores (D12-TOTAL), Borg-intensity (Borg-INT) and Borg-distress scores (Borg-DIS) at each Medical Research Council (MRC) dyspnoea grade for all three conditions combined (n = 358). Ratings for all three measures increased progressively with similar gradients of between 8% and 10% of full scale per increment in MRC grade.

Patient-reported outcomes such as the Dyspnoea-12 should be derived from the words used by patients to describe disease effects and their clinical state. We are indebted to workers who collected those descriptors in previous studies. We developed our initial 81-item pool after a systematic review of existing literature and were stuck by the similarity of items identified by different research groups. Moreover, during interview sessions patients were invited to volunteer any other terms; no new terms were provided. We are confident that the Dyspnoea-12 items are relevant, appropriate and truly capture patients’ perceptions.

The terms “work” and “effort” are often used interchangeably, in part because they tended to form into one cluster.13–15 We found that the term “more work” had more reliable patient responses than “hard work” and “more effort”. Likewise, the term “panic” is often associated with the perception of breathlessness,21 22 32 yet in this large patient cohort responses to this item were erratic, so it was removed. Two of the final 12 items did demonstrate DIF associated with diagnosis. We chose to retain these because the effect of this bias on the overall score was negligible. Furthermore, the descriptors “distressing” and “difficulty catching breath” have not previously been noted as being specific to any particular disease, so this may have been a chance finding.
Whilst we have demonstrated concurrent validity and repeatability of the Dyspnoea-12 in a different sample from that used in its development, a larger prospective study is needed to confirm these findings. The Dyspnoea-12 correlated well with Borg-distrress and HADS scores. The direction of causality in these relationships is yet to be determined, but one hypothesis is that a person’s level of psychological distress has an impact on their perceived breathlessness severity, although distress due to dyspnoea could also increase anxiety. The Dyspnoea-12 provides a method of breathlessness measurement that may allow this relationship to be explored further. It should have utility in clinical trials of new treatments targeted at the emotional impact of breathlessness, since it incorporates items related to the affective aspect of breathlessness severity. Dyspnoea-12 is currently validated for the English-speaking population; translations into other languages will require careful translation, back-translation, and cultural and linguistic validation to ensure that it performs in the same way.

In summary, the Dyspnoea-12 forms a unidimensional measure that reflects both the physical and affective aspects of dyspnoea. It addresses the need for a comprehensive dyspnoea instrument and is based on the language used by patients to describe the experience. It can measure breathlessness across several disease groups, is simple and quick to use and should find utility in routine clinical monitoring, clinical research and trials of interventions designed to ameliorate the impact of breathlessness.

Acknowledgements: The authors thank all patients and staff from Royal Brompton and Harefield NHS Foundation Trust, St George’s Hospital and North Manchester General Hospital who gave their time for this project. We are indebted to Ashi Firouzi, Professor Martin Cowie, Professor Athol Wells, Michelle Rajab, Ross Ellis, Dr Marquet Lau-Walker, Dr David Weir, Dr Georges Ng Man Kwong, Janet Mills and Sue Mason for their support with patient recruitment and administration of this project. We are also grateful to Professor Peter Barnes for his support and advice when designing this study.

Funding: The questionnaire development study was funded by the Clinical Research Committee Royal Brompton and Harefield NHS Foundation Trust. The Dyspnoea-12 validation study was funded by Action Medical Research.

Competing interests: None.

Ethics approval: The two studies performed were approved by the local ethics committees.

Provenance and peer review: Not commissioned; externally peer reviewed.

REFERENCES


APPENDIX: DYSPNOEA-12 QUESTIONNAIRE
This questionnaire is designed to help us learn more about how your breathing is troubling you. Please read each item and then tick in the box that best matches your breathing these days. If you do not experience an item tick the “none” box. Please respond to all items.

<table>
<thead>
<tr>
<th>Item</th>
<th>None</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My breath does not go in all the way</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. My breathing requires more work</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I feel short of breath</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I have difficulty catching my breath</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. I cannot get enough air</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. My breathing is uncomfortable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. My breathing is exhausting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. My breathing makes me feel depressed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. My breathing makes me feel miserable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. My breathing is distressing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. My breathing makes me agitated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. My breathing is irritating</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We do not recommend use of the instrument with data from more than three items missing. The method for scoring the Dyspnoea-12 with up to three missing items is detailed in the table below.

<table>
<thead>
<tr>
<th>Number of items missing</th>
<th>Calculation to account for missing items</th>
<th>Dyspnoea-12 score calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36 ÷ 33 = 1.09</td>
<td>1.1 × total Dyspnoea-12 score</td>
</tr>
<tr>
<td>2</td>
<td>36 ÷ 30 = 1.2</td>
<td>1.2 × total Dyspnoea-12 score</td>
</tr>
<tr>
<td>3</td>
<td>36 ÷ 27 = 1.3</td>
<td>1.3 × total Dyspnoea-12 score</td>
</tr>
</tbody>
</table>

Chronic obstructive pulmonary disease