

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

Janelle Yorke¹, Shakeeb H Moosavi², Caroline Shuldham³, Paul W Jones⁴

1. School of Nursing, Faculty of Health and Social Care, University of Salford, Greater Manchester, UK

2. National Heart and Lung Institute, Charing Cross Hospital Campus, Imperial College London, UK

3. Nursing and Clinical Governance, Royal Brompton and Harefield NHS Foundation Trust, UK

4. Division of Cardiac and Vascular Science, St George's University of London, UK

Corresponding author: Janelle Yorke, 1.43 Mary Seacole Building, School of Nursing, Faculty of Health & Social Care, University of Salford, Greater Manchester, UK. M6 6PU. j.yorke@salford.ac.uk

Word count: 2940

Key words: dyspnoea, perception, affective, outcome assessment, questionnaire

Running title : Dyspnoea-12

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in [THORAX] editions and any other BMJPG Ltd products to exploit all subsidiary rights, as set out in our licence (<http://group.bmj.com/products/journals/instructions-for-authors/licence> forms).

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. Our objective was to develop an instrument that measures overall dyspnoea severity using descriptors that reflect its different aspects.

Methods: Eighty-one 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 was 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 (χ^2 p=0.08). Items patterned into two groups called 'physical'(n=7) and 'affective'(n=5). In the separate validation study, Dyspnoea-12 correlated with Hospital Anxiety and Depression Scale (anxiety $r=0.51$; depression $r=0.44$, $p<0.001$, respectively), six-minute walk distance ($r=-0.38$, $p<0.01$), MRC Grade ($r=0.48$, $p<0.01$), and had good stability over time (intra-class 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.

Word count: 244

INTRODUCTION

Dyspnoea is a perceptual experience that is complex and highly subjective. It is a common and distressing symptom in cardiorespiratory disease. Progress has been made in identifying various aspects of dyspnoea that arise from multiple factors including environmental, psychological and physiological.[1] Other evidence indicates that dyspnoea, like pain perception, consists of 'sensory-quality' and 'affective' components,[2,3] yet no currently available dyspnoea instrument encompasses these.

Dyspnoea has been measured in two ways: (i) Directly, using modified-Borg or visual analogue scales, in persons experiencing dyspnoea at rest or in response to various stimuli;[4-6] Single item scales may not be adequate enough to capture dyspnoea complexity. (ii) Indirectly, by asking the respondent to report the level of physical activity they are not able to accomplish because of dyspnoea,[7, 8] or within scales assessing the impact of disease on quality of life.[9, 10] 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.[11,12] Studies have explored the semantics of dyspnoea, principally from a diagnostic perspective, or to understand mechanisms [13-16]. 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.[13, 16, 17] 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.[18,19]

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.[20-23] 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 (up to October 2005) using the terms: dyspnoea/dyspnea or breathless/breathlessness, and language/descriptor/questionnaire/qualitative. One patient with chronic obstructive pulmonary disease (COPD) and one with interstitial lung disease (ILD) assisted the removal of duplicate items, advised on the structure of each phrase and severity response scale used. The 81-items each with response options of 'none', 'mild', 'moderate' or 'severe', were arranged as a questionnaire list that asked patients to respond to each item that 'best matched' their current experience of being breathless. Two studies were performed, both were approved by the local ethics research committees and written consent was obtained from all participants.

STUDY ONE: Item reduction and preliminary testing

Study Participants and Measures

Breathless patients (n=358) with a primary diagnosis of COPD, ILD or chronic heart failure (CHF), and able to read English, were recruited through out-patient clinics from three hospitals in England. Each participant completed one of three randomly allocated 81-item descriptor lists, the MRC dyspnoea scale,[8] and two modified-Borg scales [6] that assessed 'current level of breathlessness' and 'current level of distress caused by breathlessness'.

Study Phases

Phase 1: Item reduction using hierarchical methods:[24] to remove the least discriminative descriptors, items were excluded if $\geq 50\%$ of patients gave a response of 'none'. Items influenced by age at $p < 0.05$ (Pearson's correlation r) were removed.

Phase 2: Item reduction using Rasch analysis: Items that survived hierarchical reduction were analysed with Rasch (RUMM2020 software).[25] Rasch models provide a template for testing how well each item contributes to the concept being measured (breathlessness severity).[26] This is an iterative process whereby the poorest fitting item is removed and the effect on the fit of the remaining is then retested. This process is continued until each item demonstrates a good fit to the model and the stability of the overall item-set to also meet the requirement for it to be a reliable unidimensional measure. In a Rasch model, severity associated with any given item is measured in 'logits' – which is the log odds of a patient of a given level of breathlessness severity, as assessed by their response to all the items combined, having a 50% chance of responding positively to that item. [27]

Individual item fit was assessed using a chi-squared statistic to compare the difference between the observed responses and those expected by the model.[28] Analysis of variance (ANOVA) was used to test each item for the influence of gender or diagnosis on the patients' response to that item, an effect known as differential item functioning (DIF). Further details regarding individual item tests of fit and DIF are provided in the e-supplement (see e-supplement figures 1-4).

The presence of any item-trait interaction was tested using a chi-square 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 (α).[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's (r) correlations examined relationships between the 12-item set total score, MRC dyspnoea grade, Borg-intensity and Borg-distress scales. The effect of diagnosis in 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 1). 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,[8] Hospital Anxiety and Depression Scale (HADS),[30] lung function tests and six-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 intra-class correlation co-efficient (ICCC). Correlations using Pearson (r) were examined between Dyspnoea-12, MRC grade, HADS, FEV₁ 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

358 patients completed the questionnaires, most (n=275, 77%) completed these during a clinic visit, the remainder for return within two-weeks. The majority were Caucasian (n=337, 94%). Their baseline characteristics are provided in Table 1.

Table 1: Study One Sample Demographics (mean±sd)

	COPD n = 123	ILD n = 129	CHF n = 106
Male gender	62 (51%)	47 (36%)	72 (68%)
Age, years	69±8	50±12	68±11
FEV ₁ %predicted	48±16	69±22	-
FVC % predicted	72±19	69±19	-
FEV ₁ :FVC % predicted	55±12	83±8	-
Left ventricular ejection fraction	-	-	35±15
MRC grade (1-5)	3.4±1.1	3.0±1.1	2.6±1.1
Borg-intensity	3.1±1.9	3.6±2.1	2.2±1.9
Borg-distress	4.5±2.4	3.7±2.2	3.4±2.4

Phase 1

Forty-seven items were initially removed because ≥50% of participants rated them as 'none', one item ('breathing more') was associated with age (p=0.02) (see e-supplement, table 1).

Phase 2

Rasch analysis was applied to the remaining 34 items. Twenty-two items were removed due to lack of fit to the Rasch model (see e-supplement, table 1). Twelve items survived the item reduction process, and make up the final item set. The 12-item set had very good internal reliability ($\alpha=0.9$). The item-trait Chi-square probability was p=0.08, supporting the conclusion that the items formed a unidimensional measure of overall breathlessness severity.

The DIF tests for two items ('distressing' and 'difficulty catching breath') suggested that they behaved differently between diseases (p<0.002) however the effect on the overall score for dyspnoea severity was negligible. Overall severity of the patients with these items included was -0.627±1.4 logits and with the items removed it was -0.625±1.4 logits, so they were retained.

Phase 3

Exploratory PCA of the 12-items identified one component with an eigenvalue >1, explaining 58% the variance (see e-supplement, table 2) and supports the conclusion from the Rasch item-trait test that the items form a unidimensional model. The second component had an eigenvalue >0.9 and explained 8% of the variance, so a two-component structure was forced to further explore the patterning of items. The first component called 'physical' contained 7 items and the second, called 'affect', contained 5 items (table 2). One item 'irritating' loaded >0.5 on both components. The apparent paradox of a unidimensional structure with perhaps two components is explained by the severity associated with the groups of items. The 'physical' items were associated with a lower overall level of dyspnoea severity (mean = -0.17 logits) whereas the 'affective' items were associated with more severe breathlessness (mean = 0.24 logits) (Figure 1).

Table 2: Principal components analysis with varimax rotation of the 12 item set.

Items were allocated to a component if they loaded >0.5, as per highlighted figures.

Item	Component 1 'Physical'	Component 2 'Affect'
My breath does not go in all the way	0.811	
My breathing requires more work	0.715	0.333
I feel short of breath	0.713	0.313
I have difficulty catching my breath	0.696	0.320
I cannot get enough air	0.688	0.398
My breathing is uncomfortable	0.653	0.456
My breathing is exhausting	0.578	0.479
My breathing makes me feel depressed		0.834
My breathing makes me feel miserable		0.821
My breathing is distressing	0.357	0.725
My breathing makes me agitated	0.488	0.658
My breathing is irritating	0.508	0.530

Phase 4

For the entire sample (n=358) the 12-item set total score correlated significantly with Borg-intensity ($r=0.47$, $p<0.001$), Borg-distress ($r=0.59$; $p<0.001$) and MRC grade ($r=0.39$; $p<0.001$). The relationship between the 12-items and MRC grade

was not dependent on diagnosis ($F=1.3$, $p >0.05$ for the interaction). The average 12-items total score, Borg-distress ratings, and Borg-intensity ratings increased progressively with similar gradients of between 8-10% of full-scale per increment in MRC grade (Figure 2).

STUDY TWO

53 patients with COPD (32 male) participated in the separate validation study. The groups mean age was 69 ± 16 ; FEV_1 (L/min) 1.43 ± 0.7 ; $FEV_1/FVC\%$ predicted $55(\pm 16)$; and 6MWD (metres) 181 ± 175 . The mean Dyspnoea-12 score for the sample was 18 ± 8 ; MRC grade 2.6 ± 1.3 ; HADS anxiety 8.7 ± 1.3 ; and HADS depression 7.7 ± 4 . Dyspnoea-12 demonstrated good internal reliability ($\alpha =0.9$) and good test-retest reliability (ICCC=0.90, $p<0.001$). Mean Dyspnoea-12 score was strongly associated with HADS scores (anxiety $r =0.51$ and depression $r=0.44$, $p<0.001$). Dyspnoea-12 correlated significantly with FEV_1 ($r= -0.30$, $p=0.03$), 6MWD ($r= -0.38$, $p<0.01$), and MRC grade ($r=0.48$, $p<0.001$).

Interview participants rated the Dyspnoea-12 as easy to complete (median=9; range=5-10), easy to understand (median=9.5; range=5-10), and helpful (median=9; range=5-10). One man with COPD commented that the questionnaire was “superb” and enabled him to say “right this is how I feel”

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 Dyspnoea-12 concurrent validity 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 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 represents the first application of Rasch methodology to the measurement of breathlessness. Our use of Rasch modeling 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 (i.e. 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 (e.g., 'makes me feel miserable') and some measure physical aspects (e.g., '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 2-components 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, not the definitive measurement of, for example the physical component severity of breathlessness.

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 struck 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 to that used in its development, a larger prospective study is needed to confirm these findings. The Dyspnoea-12 correlated well with Borg-distress 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,[32] 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, 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.

Acknowledgments: 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, Prof. Martin Cowie, Prof. Athol Wells, Michelle Rajab, Ross Ellis, Dr. Margret 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 Prof. Peter Barnes for his support and advice when designing this study.

Competing interests: None

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.

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.

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

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.

Number of items missing	Calculation to account for missing items	Dyspnoea-12 score calculation
1	$36 \div 33 = 1.09$	1.1 x total Dyspnea-12 score
2	$36 \div 30 = 1.2$	1.2 x total Dyspnea-12 score
3	$36 \div 27 = 1.3$	1.3 x total Dyspnea-12 score

References

1. American Thoracic Society. Dyspnea: Mechanisms, assessment, and management: a consensus statement. *Am J Respir Crit Care Med* 1999;**159**:321-340.
2. Wilson RC, Jones PW. Differentiation between the intensity of breathlessness and the distress it evokes in normal subjects during exercise. *Clin Sci* 1991; **80**:65-70.
3. Banzett RB, Pedersen SH, Schwartzstein RM, et al. The affective dimension of laboratory dyspnea: air hunger is more unpleasant than work/effort. *Am J Respir Crit Care Med* 2008;**177**:1384-90.
4. Fierro-Carrion G, Mahler DA, Ward J, et al. Comparison of continuous and discrete measurements of dyspnea during exercise in patients with COPD and normal subjects. *Chest* 2004; **125**:77-84.
5. Lansing RW, Moosavi SH, Banzett RB. Measurement of dyspnea: word labelled visual analogue scale vs. verbal ordinal scale. *Respir Phys Neuro* 2003;**134**:77-83.
6. Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exer* 1982;**14**:377-381.
7. Mahler DA, Ward J, Fierro-Carrion G, et al. Development of self-administered versions of modified baseline and transition dyspnea indexes in COPD. *J COPD* 2004;**1**:165-72.
8. Bestall JC, Paul EA, Garrod R, et al. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax* 1999;**54**: 581-6.
9. Jones PW, Quirk FH, Baveystock, et al. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992;**145**:1321-7.
10. Guyatt GH, Berman LB, Townsend M. Long-term outcome after respiratory rehabilitation. *CMAJ* 1987;**137**:1089-95.
11. Yorke J. Dyspnoea and pain: an interesting analogy. *J Clin Nurs*, 2008;**17**:841-2.
12. Banzett RB, Moosavi SH. Dyspnea and pain: similarities and contrasts between two very unpleasant sensations. *Am Pain Soc Bulletin* 2001;**11**: 6-8.
13. Smith J, Albert P, Bertella E, et al. Qualitative aspects of breathlessness in health and disease. *Thorax* 2009;**64**:713-718.

14. Simon PM, Schwartzstein RM, Weiss JW, et al. Distinguishable sensations of breathlessness in patients with shortness of breath. *Am Rev Respir Dis* 1990;**142**: 1009-1014.
15. Mahler DA, Harver A, Lentine T, et al. Descriptors of breathlessness in cardiorespiratory diseases. *Am J Respir Crit Care Med* 1996;**154**: 1357-63.
16. Wilcock A, Crosby V, Hughes A, et al. Descriptors of breathlessness in patients with cancer and other cardiorespiratory diseases. *J Pain Symp Man* 2002;**23**:182-189.
17. Garrard A,William M. The language of dyspnoea: a systematic review. *Internet J App Health Sci Prac* 2008;**6**: 1540-1580.
18. Parshall MB, Welsh JD, Brockopp DY, et al. Reliability and validity of dyspnea sensory quality descriptors in heart failure patients treated in an emergency department. *Heart & Lung* 2001;**30**:57-65.
19. von Leupoldt A, Balewski S, Peterson S, et al. Verbal descriptors of dyspnea in patients with COPD at different intensity levels of dyspnea. *Chest*, 2007;**132**: 141-7.
20. Caroci A, Lareau SC. Descriptors of dyspnea by patients with chronic obstructive pulmonary disease versus congestive heart failure. *Heart & Lung* 2004;**33**: 102-10.
21. Bailey PH. The dyspnea-anxiety-dyspnea cycle--COPD patients' stories of breathlessness: "It's scary /when you can't breathe". *Qual Health Res* 2004;**14**:760-78.
22. Skevington SM, Pilaar M, Routh D et al. On the language of breathlessness. *Psychol Health* 1997;**12**:677-689.
23. Michaels C, Meek PM. The language of breathing among individuals with chronic obstructive pulmonary disease. *Heart & Lung* 2004;**33**:390-400.
24. O'Leary CJ, Jones PW. The influence of decisions made by developers on health status questionnaire content. *Qual Life Res* 1998;**7**:545-50.
25. Andrich D, Lyne A, Sheridan B, et al. *RUMM2020*. RUMM Laboratory: Perth, Australia, 2003.
26. Rasch G. Probabilistic models for some intelligence and attainment tests. *Danish Institute of Educational Research* 1960.
27. Jones PW, Harding G, Berry P, et al. Development and first validation of the COPD Assessment Test. *Eur Respir J* 2009. **34**:648-654.
28. Pallant J, Tennant A. An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale. *Brit J Clin Psych* 2007. **46**: p. 1-18.
29. Kline, P. *An easy guide to factor analysis*. 1994 New York: Routledge.
30. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 1983;**67**:361-370.
31. Lansing RW, Gracely RH, Banzett RB. The multiple dimensions of dyspnea: review and hypotheses. *Respir Physio Neuro* 2009;**167**: 53-60.
32. Livermore N, Butler JE, Sharpe L, et al. Panic attacks and perception of inspiratory resistive loads in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2008;**178**:7-12.

Online Supplement

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

Janelle Yorke, Shakeeb Moosavi, Caroline Shuldham, Paul Jones

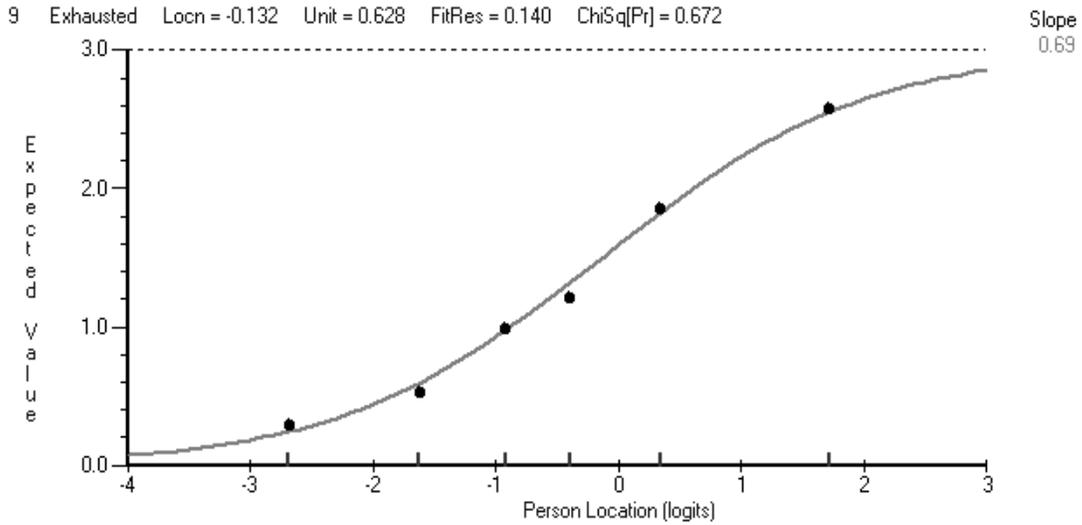
Rasch analysis

Tests of individual item fit

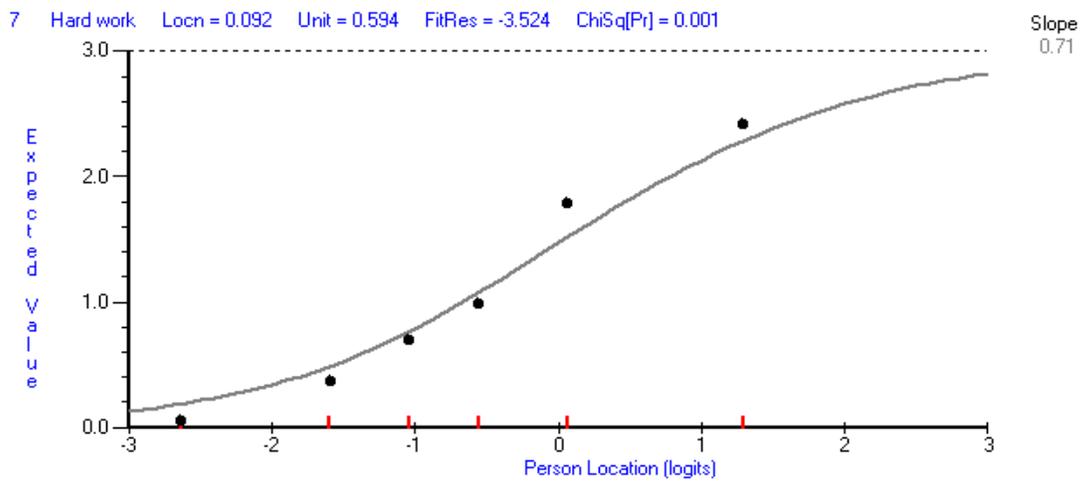
Item fit was assessed by examining the residual and Chi-squared fit statistic for each item. The item residual is a summation of the difference between the observed score and the score expected by the model for a particular item and persons. Item residuals between ± 2.5 indicate adequate fit to the model. The Chi-square (χ^2) compares the difference between the observed values with values expected by the model across different dyspnoea severity groups for each item. Severity groups, called Class Intervals, are defined by ordering all patients in terms of dyspnoea severity (determined by the responses to all items combined) and then splitting them into groups of approximately equivalent size across the sample (this is done automatically within RUMM2020). A non-significant item Chi-square ($p > 0.05$) indicates good fit to the model. Item fit is also assessed graphically using the item characteristic curve (e-supplement figures 1 and 2).

Differential item functioning

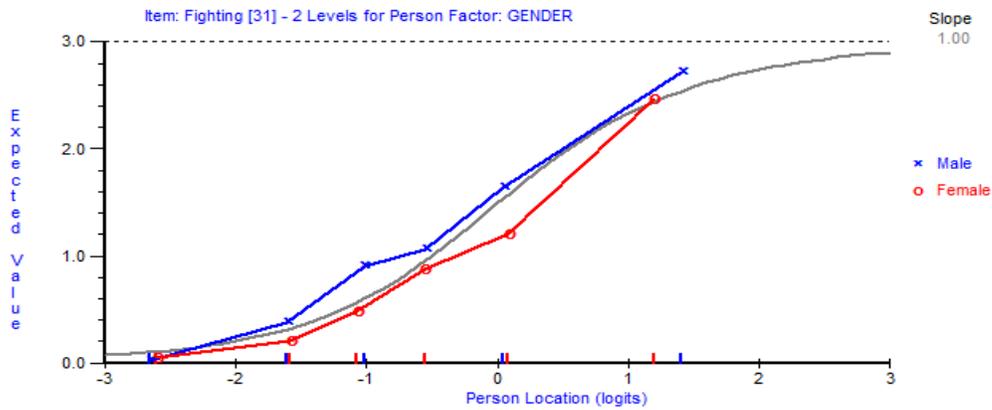
Differential item functioning (DIF) is a form of item bias and is tested using ANOVA. There are two forms of DIF. Uniform DIF is where there is uniformity in the difference of severity for an item between groups of patients. For example, where one group displays a consistently higher or lower score on a given item relative to the overall severity judged by the patients' responses to all of the items aggregated together (e-supplement figure 3). Non-uniform DIF occurs when there are differences between the groups, but with an inconsistent bias across different Class Intervals (i.e. levels of severity) (e-supplement figure 4).



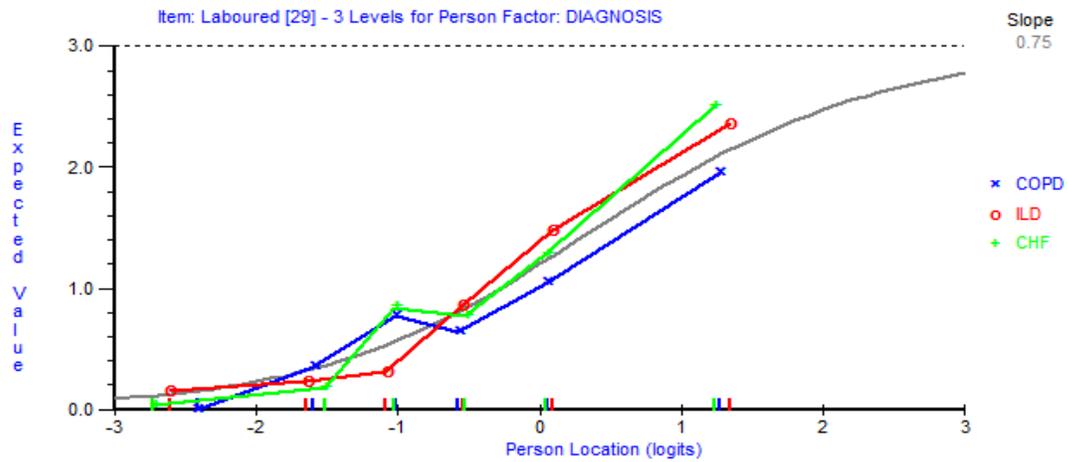
E-supplement figure 1: Item Characteristic Curve for a well-fitting item ('My breathing is exhausting'). The 'y' axis represents the item severity and the 'x' axis represents patient severity in logits. The curved line represents the expected scores for the item, and the dots represent the observed scores for the Class Intervals at the different severity levels. The fit residual (written along the top) is 0.140 and Chi-square probability is 0.672, indicating no significant deviation between the expected and observed scores for this item. This item was retained.



E-supplement figure 2: Item Characteristic Curve for the item 'My breathing is hard work'. The fit residual for this item was a high negative value (-3.524) with a significant chi-square ($p = 0.001$). This means that the item is over-discriminating - the observed scores (black dots) form a steeper curve than the expected scores (the curve). This item was removed.



E-supplement figure 3: Example of an item – ‘I am fighting for breath’ demonstrating gender associated uniform DIF. It can be seen that the female group (o line) is consistently below the male group (x line). This means that for any level of overall breathlessness severity, females had a lower (i.e. less severe) response to this particular item ($p < 0.05$). This item was removed.



E-supplement figure 4: Example of an item - 'My breathing is laboured' demonstrating diagnostic related non-uniform DIF. It can be seen that there is no organised pattern to the probability of the different disease groups affirming this item for any Class Interval level ($p < 0.05$). This item was removed.

E-supplement table 1: Details of the initial 81 items and reasons for exclusion (* denotes retained item and X denotes item removed for the reason indicated for that column)

Items	Hierarchical		Rasch analysis					
	>50% responded 'none'	Age (p<0.05)	Fit statistics		Differential item bias (p<0.05)			
			Residual ± 2.5	Chi-square (p<0.05)	Gender		Diagnosis	
					Uni	Non-u	Uni	Non-u
1. My breathing is irritating*								
2. I feel as if I am drowning	X							
3. I feel out of breath								X
4. I feel as if I am choking	X							
5. My breathing makes me feel dizzy	X							
6. I feel a raw sensation in my chest	X							
7. My breathing makes me panic			X	X				X
8. I feel as if something is stuck in my airway	X							
9. My chest feels tight			X	X				
10. I have a tight throat	X							
11. I feel weighted down	X							
12. My breath does not go in all way *								
13. My breathing requires more work *								
14. My breathing is hard work			X	X				
15. My breathing is rapid	X							
16. My breathing makes me feel restless			X	X				X
17. My voice feels tight	X							
18. I feel like I am smothering	X							
19. My breathing makes me feel lonely	X							
20. I feel a hunger for more air			X	X				X
21. My breathing is alarming	X							
22. My breathing makes me feel lightheaded	X							
23. I cannot get enough air *								
24. My breath does not go out all the way								X
25. I have a feeling of impending death	X							
26. My breathing is exhausting *								
27. My breathing makes me feel	X							

fearful						
28. I feel I need to breathe					X	
29. I have difficulty catching my breath *						
30. My breathing requires more concentration	X					
31. I feel my breath stops	X					
32. I feel that I am breathing bad air	X					
33. My breathing is tiring	X					
34. I cannot take a deep breath						X
35. I cannot breath fast enough	X					
36. My breathing makes my chest ache	X					
37. I cannot breath out fast enough	X					
38. My breathing is frightening	X					
39. My breathing is horrible	X					
40. My breathing makes me feel scared	X					
41. My breathing makes me feel like I have an itchy throat	X					
42. I feel short of breath *						
43. My breathing feels uncomfortable *						
44. My breathing makes me feel puffed			X	X		
45. My breathing is shallow			X	X		
46. I feel like I am suffocating	X					
47. My breath is heavy	X					
48. My breathing is cruel	X					
49. I have chest pain	X					
50. My breathing is too deep	X					
51. My breathing requires more effort						X
52. My breathing makes me feel agitated *						
53. My breathing is annoying						X
54. My breathing makes me feel claustrophobic	X					
55. My breathing makes me feel helplessness					X	
56. The air does not taste right	X					
57. My breathing makes me feel terrified	X					
58. I cannot control my breathing	X					
59. My breathing makes me feel anger	X					
60. I feel like my airway has become narrower					X	
61. My breathing is punishing	X					

62. I feel that I am breathing more	X	X				
63. My breathing makes me feel guilty	X					
64. My breathing makes me feel frustrated			X	X		
65. I feel wheezy			X	X		X
66. I cannot breath enough	X					
67. I feel winded in my chest	X					
68. I am gasping for breath					X	
69. My breathing feels terrible	X					
70. My breathing is distressing *						
71. My breathing is laboured					X	X
72. I am aware of my breathing						X
73. My breathing is unbearable	X					
74. I feel I need to sigh	X					
75. I feel as if I am panting	X					
76. My breathing is awful	X					
77. My chest feels constricted	X					
78. I feel I am fighting for breath					X	
79. My breathing makes me feel depressed *						
80. My breathing makes me feel miserable *						
81. I feel desperate for breath					X	

Uni: Uniform differential item functioning

Non-u: Non-uniform differential item functioning

Table 2: Initial non-rotated PCA solution

Prior to performing PCA, suitability of the data to factor analysis was assessed. The Kaiser-Meyer-Olkin value was 0.95 (exceeding the recommended value of 0.6) and Bartlett's test reached statistical significance ($p < 0.001$), supporting the suitability of the data to a PCA.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.766	56.39	58.39	6.87	58.39	58.39
2	.921	7.68	64.06			
3	.606	5.05	69.11			
4	.585	4.87	73.98			
5	.502	4.18	78.17			
6	.483	4.02	82.19			
7	.432	3.60	85.79			
8	.405	3.37	89.16			
9	.363	3.02	92.19			
10	.345	2.88	95.06			
11	.310	2.58	97.65			
12	.282	2.35	100.00			

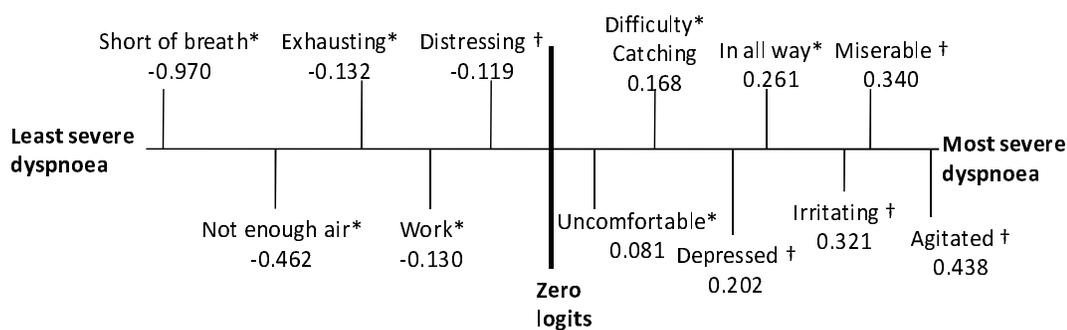


Figure 1: Illustration of logit (i.e. severity) value for each item

*Indicates item loading onto the 'physical' component identified with PCA

†Indicates item loading onto the 'affective' component identified with PCA

The values attached to each descriptor is the logit (i.e. severity) associated with that descriptor (-ve values indicates milder, +ve indicates 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.

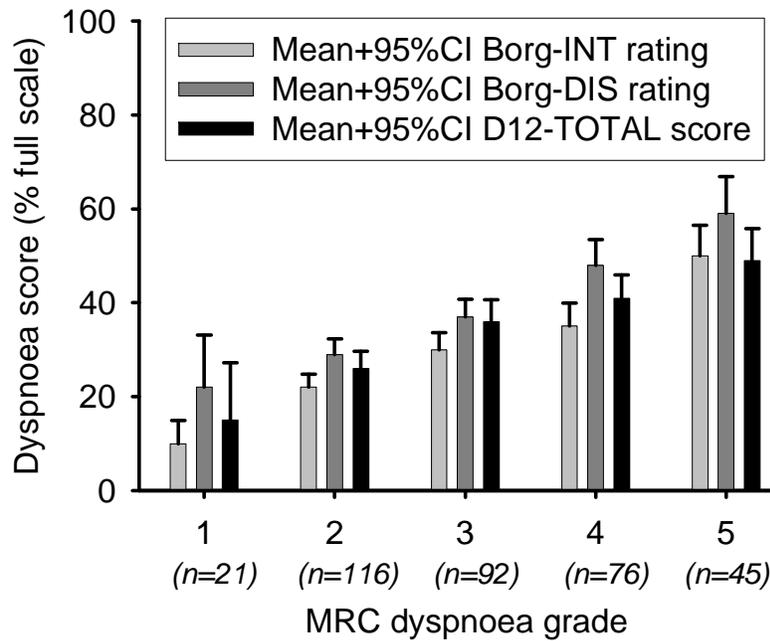


Figure 2: Mean (+95%CI) of total Dyspnea-12 scores (D12-TOTAL), Borg-intensity (Borg-INT) and Borg-distress scores (Borg-DIS) at each MRC dyspnoea grade for all three conditions combined (n=358). Ratings for all three measures increased progressively with similar gradients of between 8-10% of full-scale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