Evaluation of the Norwegian version of the Dyspnoea-12 questionnaire in patients with COPD

Andrew Malcolm Garratt [ID],[1] Ellen Margrethe Nerheim,[2] Gunnar Einvik,[3,4] Knut Stavem [ID],[3,4] Anne Edvardsen[2,4]

[1]Division for Health Services, Norwegian Institute of Public Health, Oslo, Norway
[2]LHL Hospital Gardermoen, Jessheim, Norway
[3]Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[4]Department of Pulmonary Medicine, Akershus University Hospital, Lørenskog, Norway

Correspondence to
Dr Andrew Malcolm Garratt;
andrew.garratt@fhi.no

ABSTRACT

Background The Dyspnoea-12 (D-12) questionnaire is widely used and tested in patients with breathing difficulties. The objective of this study was to translate and undertake the first evaluation of the measurement properties of the Norwegian version of the D-12 in patients with chronic obstructive pulmonary disease (COPD) attending a 4-week inpatient pulmonary rehabilitation programme.

Methods Confirmatory factor analysis was used to assess structural validity. Fit to the Rasch partial credit model and differential item functioning (DIF) were assessed in relation to age, sex and comorbidity. Based on a priori hypotheses, validity was assessed through comparisons with scores for the COPD Assessment Test (CAT), Hospital Anxiety and Depression Scales (HADS) and clinical variables.

Results There were 203 (86%) respondents with a mean age (SD) of 65.2 (9.0) years, and 49% were female. The D-12 showed satisfactory structural validity including presence of physical and affective domains. There was acceptable fit to Rasch model including unidimensionality for the two domains, and no evidence of DIF. Correlations with scores for the CAT, HADS and clinical variables were as hypothesised and highest for domains assessing similar aspects of health.

Conclusions The Norwegian version of the D-12 showed good evidence for validity and internal consistency in this group of patients with COPD, including support for two separate domains. Further testing for these measurement properties is recommended in other Norwegian patients with dyspnoea.

WHAT IS ALREADY KNOWN ON THIS TOPIC
⇒ The Dyspnoea-12 (D-12) has been widely translated and tested in patients with cardiorespiratory diseases, but a Norwegian version has not undergone testing for measurement properties.

WHAT THIS STUDY ADDS
⇒ D-12 physical and affective domains have evidence for structural validity, internal consistency, and construct validity in Norwegian patients with chronic obstructive pulmonary disease (COPD).

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE AND/OR POLICY
⇒ The D-12 is recommended for use in Norwegian patients with COPD and should be considered in other diseases with appropriate testing for measurement properties.

INTRODUCTION

Dyspnoea is one of the most distressing symptoms for people with chronic obstructive pulmonary disease (COPD) and is associated with disabling effects on quality of life and increased mortality.[1 2] In recent years, there has been an increasing focus on the multidimensional nature of dyspnoea including intensity, affective distress and impact,[2 3] which has influenced outcomes measurement.[4]

Systematic reviews have identified multiple instruments for assessing dyspnoea from the patient perspective.[5–9] Several instruments have been evaluated for measurement properties across health problems and may focus on dyspnoea severity or broader domains including physical and emotional health.[8 9] Patient reports of symptom severity often comprise single items including the modified Borg scale, Numerical Rating Scales or Visual Analogue Scales.[5] These instruments are widely used but do not assess the impact of dyspnoea on important aspects of health including emotional and functional domains. Single items are also available which assess a broader impact, including the modified Medical Research Council (mMRC) Dyspnoea Scale, which assesses disability attributable to breathlessness.[10] However, such items are often limited in terms of their measurement attributes and being limited to one domain, do not fully assess the impact of dyspnoea on health.[5 9]

The Dyspnoea-12 (D-12) assesses both physical and affective aspects of dyspnoea and was developed to provide a concise and valid measure of broad relevance across cardiorespiratory diseases.[11] Instrument content was informed by a literature review, and the resulting 81 items were reduced to 12

following consideration of item score distributions and results of Rasch analysis. D-12 items are usually summed to a single score, but six studies found evidence for two separate domains of physical and affective health.[9] D-12 scores had evidence for internal consistency and test-retest reliability in patients with COPD, interstitial lung disease and chronic heart failure.[11] The instrument has been translated into eight languages with accompanying evaluations of measurement properties.[9] Further testing has been conducted in patients with a range of health problems associated with breathing difficulties.[9 12–15] The D-12 has had widespread application as an outcome measure, and there are over 35 examples of its reported use.[9]

Following forward backwards translation, the current study assessed the Norwegian version of the D-12 in patients with COPD against recommended measurement properties including structural validity, fit to the Rasch model, internal consistency and validity through hypothesis testing.[16] The sample size in the current study permitted testing for unidimensionality and bidimensionality using confirmatory factor analysis (CFA).

METHODS

Data collection

The study included 249 potentially eligible patients with COPD aged 18 years and over attending a 4-week inpatient pulmonary rehabilitation programme at the LHL Hospital Gardermoen in South-Eastern Norway over a 6-month period from June 2018. Patients were included if they had no cognitive impairments and sufficient understanding of the Norwegian language. The self-completed pen and paper questionnaire which included the D-12, were completed within the first week of attendance. Patients received the questionnaire in a group setting for completion at the end of the session in the same room or in their own room. There was a nurse available to answer any questions during completion and on collection from patients who completed it in their room.

Patient-reported outcomes

The D-12 has 12 items with four-point scales of none, mild, moderate and severe in relation to breathlessness 'these days'.[11] The first seven items sum to the physical domain with scores from 0 to 21. The remaining five items sum to the affective domain scores from 0 to 15. All twelve items sum to the D-12 total scores from 0 to 36. Higher scores represent greater severity.[11] Translation of the original English version of the D-12 into Norwegian followed international recommendations for forward-backwards translation of patient-reported outcome measures (PROMs)[17 18] including two independent forward and backwards translations. The original time frame of 'these days' was retained following translation. Minor discrepancies between translations were resolved by discussion within the research group which included the translators. The Norwegian translation (online

supplemental figure S1), including instructions, questions and scaling, was found to be acceptable and easy to understand following interviews with 20 patients with COPD who did not participate in the main study.

In addition to the D-12, the questionnaire included the COPD Assessment Test (CAT) and Hospital Anxiety and Depression Scale (HADS). The CAT has eight items with six-point scales with endpoint-only descriptors.[19] Four items relate to symptoms and four relate to broader aspects of health and quality of life. The items sum to give a score from 0 to 40, where higher scores represent a greater impact of COPD on health. The CAT has been widely translated and has evidence for reliability, validity and responsiveness.[19 20] The HADS was developed to assess mood disorder in non-psychiatric hospital outpatients and has 14 items with four-point descriptive scales that refer to the last week.[21] Seven items sum to give the anxiety scale and the remainder sum to give the depression scale. Scores range from 0 to 21 with higher scores representing more symptoms. The HADS has been widely used alongside the D-12[9] and has evidence for internal consistency and validity in Norwegian patients with COPD.[22]

Clinical measures

Dyspnoea was assessed by the mMRC Dyspnoea Scale, as registered by physicians in the electronic medical record. The single-item mMRC assesses disability attributable to shortness of breath and has been widely used in Norwegian healthy and patient populations.[10] Spirometry was conducted according to American Thoracic Society/European Respiratory Society recommendations using appropriate reference values.[23] A 6 min walk test followed standard recommendations reporting walked distance in metres.[24] Dyspnoea before and after the walk test was assessed using the Borg CR10 scale with a range of 0–10.[25] Further information retrieved from medical records included age, sex, smoking status, employment status, body mass index, number of COPD-related hospitalisation in the last year, whether there was a current acute exacerbation, use of long-term oxygen treatment and comorbidities.

Statistical analysis

Descriptive statistics are presented with the mean (SD), median (range) or number (%), as appropriate. Missing data and floor and ceiling effects were assessed at the item and domain level. CFA with robust-weighted least squares (WLSMV)[26] was used to assess the structural validity of the D-12, or the extent to which the item scores adequately contribute to physical, affective and overall domains.[16] Model fit was assessed by the root mean square error approximation (RMSEA, acceptable fit if <0.06), the Comparative Fit Index (CFI, acceptable fit if >0.95, poor fit if <0.90, otherwise marginal) and the Tucker Lewis Index (TLI, acceptable fit if >0.95, poor fit if <0.90, otherwise marginal).[16 27]

Table 1 Background characteristics of patients completing the questionnaire (n=203)

	Mean (SD)	n (%)
Female		100 (49.3)
Age years	65.2 (9.0)	
Current smoker		48 (23.8)
Employment status:		
Receiving disability benefits		110 (54.2)
Retired		58 (28.7)
Sick leave		19 (9.4)
In work		29 (14.4)
mMRC grade (0–4)		
0		–
1		30 (14.9)
2		60 (29.7)
3		62 (30.7)
4		50 (24.8)
Spirometry:		
FEV_1 (postbronchodilator, % predicted)	42 (19)	
FEV_1/FVC (postbronchodilator)	0.45 (0.13)	
6 min walking test distance (metres)	373 (129)	
Borg CR10 scale (range 0–10):		
Pre-6MWT	1.7 (1.2)	
Post-6MWT	6.1 (2.1)	
Body mass index (BMI)*, kg/m^2	25.2 (5.7)	
Obesity (BMI >30.0 kg/m2)		36 (17.7)
No of hospitalisations for COPD last year		
0		107 (52.7)
1		42 (21.0)
2		20 (10.0)
≥3		31 (15.5)
Current acute exacerbation n (%)		50 (24.6)
Long-term oxygen treatment n (%)		22 (10.9)
Comorbidity†:		
Asthma		46 (22.7)
Coronary heart disease		29 (14.4)
Diabetes		16 (7.9)
Heart failure		14 (7.0)
Hypertension		55 (27.1)
Obstructive sleep apnoea		20 (10.0)

*Height and weight measured under rehabilitation.
†Derived from medical records.
COPD, chronic obstructive pulmonary disease; CR10, Category-Ratio 10; FEV, forced expiratory volume; FVC, Forced Vital Capacity; mMRC, modified Medical Research Council; MWT, Minute Walking Test.

Unidimensionality of D-12 domains was tested using the Rasch partial credit model. This extension of the Rasch model for polytomous items has separable item and person parameters, sufficient statistics and conjoint additivity allowing item and person comparisons.[28] Overall and item fit statistics were used to assess whether items within the domains fitted the one-dimensional model. Item fit was assessed with the χ^2 statistic, standardised residuals, which should be between ±2.5, and item characteristic curves. Local independence was assessed through examination of the residual correlation matrix, coefficients of ≥0.2 indicating item redundancy.[29 30]

Differential item functioning (DIF) occurs when different groups with the same levels of construct being measured, respond differently to an item. Consistent differences across the construct represent uniform DIF. Inconsistent differences that vary across the construct represent non-uniform DIF. Both forms of DIF were assessed for gender, pensionable age (67 years), receipt of disability payment and four comorbidities: anxiety/depression, asthma, hypertension and obesity. Differences of ≥0.5 logits in item difficulties were considered meaningful in interpretation of DIF.[31 32]

Internal consistency was assessed by the Person Separation Index (PSI)[33] and Cronbach's alpha.[16] Estimates of 0.70 and 0.90 are necessary for group and individual comparisons respectively.[16]

Hypothesis testing was used to further assess the validity of D-12 scores through comparisons with those for the HADS and clinical measures.[9–15] Criteria for expected levels of correlation followed those from a systematic review of PROMs.[34] First, correlations ≥0.60 were expected for scores assessing the same construct: D-12 affective and HADS anxiety and depression. Second, correlations <0.60 and ≥0.30 were expected for scores assessing largely related but dissimilar constructs: D-12 domains/total and CAT. Third, correlations <0.50 and ≥0.20 were expected for moderately related but dissimilar constructs: D-12 physical/total and HADS, mMRC, Borg scores, FEV_1, number of COPD hospitalisations in the last year and 6 min walk distance. Fourth, correlations <0.30 were expected for weakly related or unrelated constructs: D-12 affective and mMRC, Borg scores, FEV_1, number of COPD hospitalisations in the last year and 6 min walk distance. At least 75% of the results should correspond with the hypotheses.[16 34] Pearson's correlation was used to aid comparisons with published studies, and findings are shown alongside those from a systematic review.[9]

Statistical analyses were undertaken using RUMM2020 V.4.1 (Rumm Laboratory, Perth, Western Australia), Lisrel version 7 (Muthe'n & Muthe'n, Los Angeles, CA) and Stata V.15.0 (StataCorp).

Patient and public involvement

Patients and public were not involved in development of the research question, study design, study conduct or dissemination.

Table 2 Descriptives for Dyspnoea-12 items and domains (n=203)

Scale/item	Missing %	Mean (SD) (scale)/ median (item)*	None %	Mild %	Moderate %	Severe %
Physical†	0	10.9 (5.3)	3.0	–	–	3.4
My breath does not go all the way in	0.5	1	17.2	33.5	36.0	13.3
My breathing requires more work	1.0	2	13.4	31.7	43.1	11.9
I feel short of breath	1.0	2	5.9	16.8	46.5	30.7
I have difficulty catching my breath	1.5	1	18.4	33.8	35.3	12.4
I cannot get enough air	1.0	2	10.4	25.7	39.1	24.8
My breathing is uncomfortable	1.5	1	27.4	31.3	31.8	9.5
My breathing is exhausting	2.0	1	21.0	31.5	32.0	15.5
Affective†	0.5	7.2 (4.5)	5.9	–	–	7.4
My breathing makes me feel depressed	0.5	1	26.6	31.0	23.2	19.2
My breathing makes me feel miserable	0.5	1	26.1	30.0	25.6	18.2
My breathing is distressing	0.5	2	19.7	26.6	33.0	20.7
My breathing makes me agitated	1.5	1	28.9	26.9	33.3	10.9
My breathing is irritating	1.0	2	13.4	26.7	36.1	23.8

*Item scores range from 0 to 3. None and severe correspond with the lowest and highest scores, respectively.
†Physical and affective domain scores range from 0 to 21 and from 0 to 15, respectively. Higher scores represent greater severity.

RESULTS

Data collection

Of the 249 patients assessed for eligibility, 12 did not meet inclusion criteria, 32 did not want to participate and 203 (86%) consenting participants completed the questionnaire. Table 1 shows their background and clinical characteristics. The mean age (SD) was 65.2 (9.0), 49% were female, and 54% were receiving disability benefits.

Distribution of scores

Levels of missing data were low for all D-12 items (table 2). Median item scores corresponded with the mild and moderate response categories for seven and five items, respectively. Item scores were approximately normally distributed, except for the affective domain item relating to agitation, which had a slightly lower number of responses for the mild compared with adjacent response categories. For physical domain items, floor effects ranged from 6% to 27% for "I feel short of breath" and "My breathing is uncomfortable", respectively. For affective domain items, floor effects ranged from 13% to 29% for "My breathing is irritating" and "My breathing makes me agitated", respectively. For physical domain items, ceiling effects ranged from 10% to 31% for "My breathing is uncomfortable" and "I feel short of breath", respectively. For affective domain items, ceiling effects ranged from 11% to 21% for "My breathing makes me agitated" and "My breathing is irritating", respectively.

Statistical analysis

CFA results showed good model fit for the D-12 physical and affective domains according to all fit indices (table 3). The D-12 unidimensional model had poor fit according to the RMSEA and TLI, and marginal fit according to the CFI. The χ^2 test also showed that the bidimensional model had better model fit. The two domain scores had a correlation of 0.57.

After adjusting for missing values based on recommendations,[11] there was complete data for the physical domain and one missing response for the affective domain scores. Both mean domain scores were approximately in the middle of the possible score ranges with floor and ceiling effects of between 3% and 7% (table 2).

Table 3 Confirmatory factor analysis goodness of fit indices for the Dyspnoea-12

Model*	χ^2	df	Comparative Fit Index†	Tucker-Lewis Index†	Root mean square error of approximation‡
Bidimensional	52.68	53	1.00	1.00	0.008
Unidimensional	369.68	54	0.91	0.89	0.175

χ^2 test of model fit was non-significant (p<0.05) for the bidimensional model and significant (p<0.001) for the unidimensional model.
*Bidimensional model with physical (items 1–7) and affective (items 8–12) domains. Unidimensional with a single overall domain.
†Acceptable fit if >0.95, poor fit if <0.90, otherwise marginal.
‡Acceptable fit if <0.06.

Table 4 Rasch analysis for the Dyspnoea-12 (n=203)

Scale/item (overall fit p value*)	Location†	Fit residual	χ2	P value	PSI‡, Cronbach's alpha
Physical function (0.85)					0.89, 0.92
My breath does not go all the way in	−0.32	1.29	2.50	0.29	
My breathing requires more work	0.15	−1.56	3.73	0.16	
I feel short of breath	−1.38	0.62	0.07	0.97	
I have difficulty catching my breath	0.43	−0.38	0.97	0.62	
I cannot get enough air	−0.75	0.26	0.25	0.88	
My breathing is uncomfortable	0.90	−0.87	0.71	0.70	
My breathing is exhausting	0.33	0.40	0.55	0.76	
Affective (0.31)					0.89, 0.93
My breathing makes me feel depressed	0.17	−1.84	2.74	0.25	
My breathing makes me feel miserable	0.22	−2.15	3.93	0.14	
My breathing is distressing	−0.31	0.40	0.30	0.86	
My breathing makes me agitated	0.84	−0.62	0.31	0.86	
My breathing is irritating	−0.91	3.15	4.35	0.11	

*Overall fit p- value for χ2, where a non-significance (p>0.05) indicates fit to the Rasch model.
†Location is the item position on the latent scale or level of health assessed. Fit residuals are the difference between the observed and scores for the item and χ2 where a non-significance (p>0.05) indicates fit to the Rasch model.
‡Person Separation Index is an estimate of reliability or the proportion of error free variance of the distribution of person estimates relative to the sum of this variance and the error variance in these estimates.

Table 4 shows that items comprising the D-12 physical and affective domains fit the Rasch model according to the p values for the χ^2 statistics. No items had disordered thresholds and there was no evidence for DIF. Item residual correlations did suggest a lack of local independence for first two items within the affective domain, "My breathing makes me depressed" and "My breathing makes me miserable". These two items were summed to a single item and satisfactory fit to the Rasch model was found including item fit residuals between ±2.5. Levels of Cronbach's alpha met the criterion for group and individual comparisons (table 4). PSI levels were borderline for the more stringent criterion relating to individual comparisons.

Correlations between D-12 scores and those for the other instruments and clinical variables were highly consistent with a priori hypotheses (table 5). The highest levels were found for D-12 affective domain scores and those for the HADS, followed by those for the CAT. Lower levels of correlation were found for mMRC, Borg scores and remaining clinical measures. Of the 27 correlations, 21 (78%) were within the hypothesised range. Two exceptions related to HADS scores. First, correlation between the D-12 affective domain scores and those for HADS depression fell just below the criterion of 0.6. Second, correlation between the D-12 total scores and HADS anxiety were above the criterion of 0.6. The remaining exceptions were the four correlations between D-12 physical and total scores and those for the Borg before the walking test and FEV$_1$ predicted. These were below the criterion of 0.2.

DISCUSSION

The Norwegian D-12 performed well in relation to measurement criteria widely recommended in the evaluation of PROMs[16] and further support it as an appropriate measure of dyspnoea across populations, settings and languages.[9]

Levels of missing data were low across the 12 items, and the distribution of item scores was approximately normal. CFA results showed that the Norwegian D-12 bidimensional model had good evidence for structural validity, including the presence of the physical and affective domains. These findings concur with six other studies.[9] Given the widespread use of the D-12 score based on all 12 items, the current study also tested empirical support for the unidimensional model. Model fit was poor according to widely recommended criteria.[16] This evidence suggests focus should be on physical and affective domain scores, and that summed scores should be interpreted alongside the two domains. Furthermore, this concords with the multidimensional model proposed.[9 11]

This is first study to assess the fit of both D-12 domains to the Rasch model and the results were satisfactory. Unidimensionality of the two domains was further confirmed, and both had acceptable levels of internal consistency, close to or meeting the more stringent criterion of 0.9.[16] The levels of Cronbach's alpha were highly comparable to those found across seven studies, which had mean levels of 0.91 and 0.93 for the physical and affective domains, respectively.[9] Hence, the domain scores are suitable for group and individual level comparisons.

Table 5 Correlations* (hypotheses) and review findings[9] for Dyspnoea-12 scores, patient-reported data and clinical measures (n=203)

| | Dyspnoea-12 | | | | | |
| | Physical | | Affective | | Total scores | |
	r, hypothesised	Review mean, range	r, hypothesised	Review mean, range	r, hypothesised	Review mean, range
CAT	0.48	0.62	0.45	0.53	0.53	0.54
	(0.3≤r < 0.6)	–	0.3≤r < 0.6	–	0.3≤r < 0.6	0.35–0.67
HADS						
Anxiety	0.44	0.45	0.74	0.61	0.65	0.54
	(0.2≤r <0.5)	0.33–0.49	r≥0.6	0.54–0.71	0.3≤r < 0.6	0.35–0.64
Depression	0.42	0.41	0.57	0.48	0.55	0.47
	(0.2≤r <0.5)	0.22–0.62	r≥0.6	0.24–0.68	0.3≤r < 0.6	0.22–0.68
mMRC†	0.36	0.55	0.28	0.41	0.36	0.41
	(0.2≤r <0.5)	0.48–0.64	r<0.3	0.22–0.52	0.2≤r <0.5	0.39–0.60
Borg CR10						
Before 6MWT	0.17^{NS}	–	0.13^{NSr}	–	0.17 ^{NS}	–
	(0.2≤r <0.5)	–	r<0.3	–	0.2≤r <0.5	–
After 6MWT	0.26	–	0.13^{NS}	–	0.22	–
	(0.2≤r <0.5)	–	r<0.3	–	0.2≤r <0.5	–
FEV$_1$ predicted‡	0.19	0.22	0.06^{NS}	0.16	0.14 ^{NS}	0.26
	(0.2≤r <0.5)	0.06–0.35	r<0.3	0.12–0.23	0.2≤r <0.5	0.08–0.38
Hospitalisations§	0.24	–	0.28	–	0.29	–
	(0.2≤r <0.5)	–	r<0.3	–	0.2≤r <0.5	–
6MWT distance¶	0.32	0.35	0.26	0.22	0.33	0.35
	(0.2≤r <0.5)	0.16–0.52	r<0.3	0.08–0.49	0.2≤r <0.5	0.13–0.53

*Pearson correlations reported in the same direction for ease of interpretation. Statistically significant (p<0.01) unless stated: NS not significant.
†modified Medical Research Council dyspnoea score.
‡Forced expiratory volume in 1 s in percentage of predicted value.
§Number of hospitalisations in the last year.
¶6 min walking test.
CAT, COPD Assessment Test; CR10, Category-Ratio 10; HADS, Hospital Anxiety and Depression Scale; mMRC, modified Medical Research Council.

D-12 items were not affected by DIF for sex, age and comorbidity, but there was some evidence for a lack of local independence for two items from the affective domain. This finding suggests that responses to these items are dependent on one another. Combining the items gave satisfactory fit to the Rasch model, but together with the issues of bidimensionality versus unidimensionality, this should be explored in other patient populations including international studies.

Validity testing was strengthened through the inclusion of PROMs and clinical variables that have been widely used alongside the D-12,[9 11–15] together with a priori hypotheses. The levels of correlation found between the physical domain and HADS scores were approximately in the middle of the ranges across six published studies: 0.33–0.49 and 0.22–0.62 for symptoms of anxiety and depression, respectively.[9] The correlation between the D-12 affective domain and HADS depression scores was slightly lower than expected, but the content of the former relates more to symptoms of anxiety. Correlations for the affective domain were slightly above or nearer the upper ranges reported for the same six studies: 0.54–0.71 and 0.24–0.68 for anxiety and depression, respectively.[9] There were similar findings for the D-12 total scores relative to those from previous studies.[9]

One published study correlated D-12 domain scores with those for the CAT.[9 35] Somewhat lower correlations were found in the current study for both the physical (0.48 vs 0.62) and affective (0.45 vs 0.53) domains.[35] The level of correlation found for the total scores and CAT was very close to the mean level found across five studies (0.54 vs 0.53). The mean correlation for the three published studies that included the mMRC was higher for physical (0.36 vs 0.55), affective (0.28 vs 0.41) and D-12 total scores (0.36 vs 0.41).[9] The lower levels of correlation found in the current study might be attributable to

physician completion rather than patient completion of the mMRC in earlier studies.

The current findings concur with those of the systematic review and lower correlations between D-12 scores and those for clinical variables including the FEV_1 and 6 min walk distance.[9] While several of the correlations were lower than previous findings,[9] the majority of correlations between D-12 scores and those for other questionnaires and clinical variables were as hypothesised, and hence, met the criterion of 75%.[16] Except for CAT scores, correlations were generally slightly higher for D-12 domain scores compared with total scores, and particularly in relation to the HADS. The findings concur with the systematic review[9] and taken together, are further support for the bidimensional nature of the D-12.

Strengths and limitations

The study was broadly similar in scope to eight international studies reporting on the translation and testing of the D-12.[9] The sample size was adequate for the application of CFA and the Rasch partial credit model. Most studies reporting on the D-12 had sample sizes of well under 100,[9] which limited more advanced testing for measurement properties, including structural validity. This may be why the D-12 continues to be scored as a unidimensional instrument. Again, facilitated by sample sizes, this was the first time the Rasch partial credit model was used to assess the two separate domains with evidence for fit to the model. The current study included additional PROMs and clinical variables which have been widely used in testing of the D-12.[9] Testing followed international recommendations and was based on explicit hypotheses.[16]

The study did not include a test–retest design which would have provided information on the stability of scores together with the SE of measurement and minimal important differences.[9 16] Two studies have reported test–retest reliability for domain scores with similarly high levels of reliability and internal consistency found with UK asthma patients,[12] while somewhat lower levels were found for a mixed population in Sweden.[35] The cross-sectional nature of the study also precluded testing for responsiveness to changes in health. Finally, this study was limited to patients with COPD, and further testing in other Norwegian populations is necessary before the D-12 can be more widely recommended.

CONCLUSIONS

In conclusion, the Norwegian-language D-12 showed acceptable evidence for measurement properties including structural validity, internal consistency and construct validity, in a relatively large sample of patients with COPD when compared with existing studies. Testing for structural validity supported the two domains of physical and affective health. Applications of the D-12 should focus on domain rather than total scores, but further testing for structural validity is recommended in different populations. Evidence supporting the responsiveness of the D-12 should be investigated in future studies.

ORCID iDs
Andrew Malcolm Garratt http://orcid.org/0000-0002-1000-4780
Knut Stavem http://orcid.org/0000-0003-4512-8000

REFERENCES

1 Celli BR, Cote CG, Marin JM, *et al*. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med* 2004;350:1005–12.
2 Parshall MB, Schwartzstein RM, Adams L, *et al*. An official American thoracic Society statement: update on the mechanisms, assessment, and management of dyspnea. *Am J Respir Crit Care Med* 2012;185:435–52.
3 Laviolette L, Laveneziana P, ERS Research Seminar Faculty. Dyspnoea: a multidimensional and multidisciplinary approach. *Eur Respir J* 2014;43:1750–62.
4 Banzett RB, O'Donnell CR, Guilfoyle TE, *et al*. Multidimensional dyspnea profile: an instrument for clinical and laboratory research. *Eur Respir J* 2015;45:1681–91.
5 Bausewein C, Farquhar M, Booth S, *et al*. Measurement of breathlessness in advanced disease: a systematic review. *Respir Med* 2007;101:399–410.
6 Liu Y, Li H, Ding N, *et al*. Functional status assessment of patients with COPD: a systematic review of performance-based measures and patient-reported measures. *Medicine* 2016;95:e3672.
7 Tsiligianni I, Kocks JWH. Daytime symptoms of chronic obstructive pulmonary disease: a systematic review. *NPJ Prim Care Respir Med* 2020;30:6.
8 Tinti S, Parati M, De Maria B, *et al*. Multi-Dimensional Dyspnea-Related scales validated in individuals with cardio-respiratory and cancer diseases. A systematic review of psychometric properties. *J Pain Symptom Manage* 2022;63:e46–58.

9 Williams MT, Lewthwaite H, Paquet C, *et al*. Dyspnoea-12 and multidimensional dyspnea profile: systematic review of use and properties. *J Pain Symptom Manage* 2022;63:e75–87.
10 Grønseth R, Vollmer WM, Hardie JA, *et al*. Predictors of dyspnoea prevalence: results from the BOLD study. *Eur Respir J* 2014;43:1610–20.
11 Yorke J, Moosavi SH, Shuldham C, *et al*. Quantification of dyspnoea using descriptors: development and initial testing of the Dyspnoea-12. *Thorax* 2010;65:21–6.
12 Yorke J, Russell A-M, Swigris J, *et al*. Assessment of dyspnea in asthma: validation of the Dyspnea-12. *J Asthma* 2011;48:602–8.
13 Swigris JJ, Yorke J, Sprunger DB, *et al*. Assessing dyspnea and its impact on patients with connective tissue disease-related interstitial lung disease. *Respir Med* 2010;104:1350–5.
14 Lee BY, Lee S, Lee JS, *et al*. Validity and reliability of cat and Dyspnea-12 in bronchiectasis and tuberculous destroyed lung. *Tuberc Respir Dis* 2012;72:467–74.
15 Tan J-Y, Yorke J, Harle A, *et al*. Assessment of breathlessness in lung cancer: psychometric properties of the Dyspnea-12 questionnaire. *J Pain Symptom Manage* 2017;53:208–15.
16 Prinsen CAC, Mokkink LB, Bouter LM, *et al*. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.
17 Wild D, Grove A, Martin M, *et al*. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (pro) measures: report of the ISPOR Task force for translation and cultural adaptation. *Value Health* 2005;8:94–104.
18 Mokkink LB, Prinsen CA, Patrick DL. COSMIN study design checklist for patient-reported outcome measurement instruments, 2019. Available: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf [Accessed 22 Feb 2022].
19 Karloh M, Fleig Mayer A, Maurici R, *et al*. The COPD assessment test: what do we know so far?: a systematic review and meta-analysis about clinical outcomes prediction and classification of patients into gold stages. *Chest* 2016;149:413–25.
20 Zhou Z, Zhou A, Zhao Y, *et al*. Evaluating the clinical COPD questionnaire: a systematic review. *Respirology* 2017;22:251–62.
21 Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
22 Bratås O, Grønning K, Forbord T. Psychometric properties of the hospital anxiety and depression scale and the general health Questionnaire-20 in COPD inpatients. *Scand J Caring Sci* 2014;28:413–20.
23 Quanjer PH, Tammeling GJ, Cotes JE, *et al*. Lung volumes and forced ventilatory flows. Report Working Party standardization of lung function tests, European community for steel and coal. official statement of the European respiratory Society. *Eur Respir J Suppl* 1993;16:5–40.
24 Holland AE, Spruit MA, Troosters T, *et al*. An official European respiratory Society/American thoracic Society technical standard: field walking tests in chronic respiratory disease. *Eur Respir J* 2014;44:1428–46.
25 Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc* 1982;14:377–81.
26 Brown TA. *Confirmatory factor analysis for applied research*. New York: The Guilford Press, 2006.
27 Hu Li-tze, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 1999;6:1–55.
28 Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982;47:149–74.
29 Baghaei P. Local dependency and Rasch measures. *Rasch Measure Transact* 2008;21:1105–6.
30 Christensen KB, Makransky G, Horton M. Critical Values for Yen's Q_3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Appl Psychol Meas* 2017;41:178–94.
31 Teresi JA, Fleishman JA. Differential item functioning and health assessment. *Qual Life Res* 2007;16 Suppl 1:33–42.
32 Rouquette A, Hardouin J-B, Vanhaesebrouck A, *et al*. Differential item functioning (DIF) in composite health measurement scale: recommendations for characterizing DIF with meaningful consequences within the Rasch model framework. *PLoS One* 2019;14:e0215073.
33 Wright BD, Masters GN. *Rating scale analysis*. Chicago: MESA Press, 1982.
34 Chiarotto A, Terwee CB, Kamper SJ, *et al*. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain: a systematic review. *J Clin Epidemiol* 2018;102:23–37.
35 Sundh J, Bornefalk H, Sköld CM, *et al*. Clinical validation of the Swedish version of Dyspnoea-12 instrument in outpatients with cardiorespiratory disease. *BMJ Open Respir Res* 2019;6:e000418.