International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 17(2): 104–110 (2008)
Published online 18 March 2008 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/mpr.234

Examining item bias in the anxiety subscale of the Hospital Anxiety and Depression Scale in patients with chronic obstructive pulmonary disease

WAI-KWONG TANG,[1] ERIC WONG,[2] HELEN F.K. CHIU,[1] C.M. LUM,[3] GABOR S. UNGVARI[1]

1 Department of Psychiatry, Chinese University of Hong Kong, Hong Kong, SAR, China
2 Centre for Epidemiology and Biostatistics and School of Public Health, Chinese University of Hong Kong, Hong Kong, SAR, China
3 Department of Medicine and Geriatrics, Shatin Hospital, Hong Kong, SAR, China

Abstract

The Hospital Anxiety and Depression Scale (HADS) is a widely used screening instrument for depression and anxiety in medically compromised patients. The purpose of this study was to examine the differential item functioning (DIF) of the anxiety subscale of the HADA (HADS-A). A research assistant administered the HADS-A to 166 Chinese patients with chronic obstructive pulmonary disease (COPD) who were consecutively admitted to a rehabilitation hospital. Although the HADS-A was overall uni-dimensional, there were one mute item and two items with borderline misfit. Only one item had a DIF for arterial oxygen saturation. No item had DIF for other indicators of the severity of COPD. In conclusion, this study found that for one item the HADS-A has significant item bias for the severity of disease in patients with COPD. Copyright © 2008 John Wiley & Sons, Ltd.

Key words: Hospital Anxiety and Depression Scale (HADS), anxiety, screening, Rasch analysis, chronic obstructive pulmonary disease

Introduction

The Hospital Anxiety Depression Scale (HADS) (Zigmond and Snaith, 1983) performs well in assessing the symptom severity and caseness of anxiety disorders and depression in both medically ill, psychiatric and primary care patients and in the general population (Herrman, 1997; Bjelland et al., 2002).

There is evidence suggesting that demographic variables, such as age (Herrman, 1997; Hinz and Schwarz, 2001), sex (Herrman, 1997; Hinz et al., 2002) and education (Osborne et al., 2004) influence the HADS anxiety score (HADS-A). These demographic variables may indeed influence the degree of anxiety. However, the variation of HADS-A scores in different patient groups may also be explained in terms of differential item functioning (DIF) which refers to a systematically higher or lower response to one or more items in a group, independent of the trait that is supposed to be measured. Item response theory stipulates that given the same trait level the probability of affirming an item should be the same between groups (Scheuneman, 1979), or the item should not have significant DIF for any subgroup of patients.

DIF by culture (Azocar et al., 2001), age (Crane et al., 2004), education (Jones and Gallo, 2002), sex (Smith and Reise, 1998), mode of administration (Chan et al., 2004) or the nature of the target illness (Sprangers and Schwartz, 1999) exists in various

Copyright © 2008 John Wiley & Sons, Ltd

psychiatric scales. Examining of DIF is important as it may affect the interpretation of cross-cultural study (Azocar et al., 2001) as well as studies comparing different age groups (Crane et al., 2004).

HADS is frequently applied in chronic obstructive pulmonary disease (COPD) research (Dowson et al., 2001; Oga et al., 2002; Garuti et al., 2003). COPD is associated with an increased prevalence of anxiety disorders (Smoller et al., 1996). The symptoms of panic attacks and pulmonary disease overlap, so that panic anxiety can mimic cardiopulmonary diseases and, conversely, dyspnea can masquerade as anxiety disorder (Smoller et al., 1996). According to the suffocation false alarm model (Klein, 1993), panic attacks could be triggered by a hypersensitive brainstem autonomic control mechanism that can initiate both panic and a sensation of dyspnea. It is not known whether the severity of the COPD affects the performance of HADS-A.

The aim of the present study was to examine the feasibility of using the HADS-A in a hospitalized sample of Hong Kong Chinese patients with COPD by applying Rasch analysis. Specifically, we set out to determine the DIF for the severity of the underlying lung disease of the HADS-A.

Method

Study population

Patients of Chinese ethnicity were recruited from the inpatient pulmonary rehabilitation unit (PRU) in Shatin Hospital, a rehabilitation center for a catchment area of a population of 1.2 million. The diagnosis of COPD was made by a pulmonary physician on the basis of history, physical examination and spirometry data, if available [forced expiratory volume in one second percent predicted (FEV_1%) <75% of predicted value]. The exclusion criteria were physical frailty precluding assessment, a Mini Mental State Examination (MMSE; Chiu et al., 1994) score <15 or inability to give consent.

The study protocol was approved by the Clinical Research Ethics Committee of the Faculty of Medicine, Chinese University of Hong Kong. All subjects gave written informed consent.

Measurements

The following demographic and medical data were collected: sex, age, marital status, education, occupation, duration of COPD, arterial oxygen saturation (pO_2), six-minute walking distance (Guyatt et al., 1985), Forced vital capacity percent predicted (FVC%) and forced expiratory volume in one second percent predicted (FEV_1%). A research assistant administered the Chinese version of the HADS-A (Leung et al., 1999) to all consecutively admitted patients who met entry criteria of the study one week after their admission to PRU. The questions were read aloud to all patients rather than being self-administered to overcome difficulties like poor vision and illiteracy. The score of each item on the HADS-A ranged from 0 to 3; higher scores indicate more severe anxiety symptoms.

Statistical analyses

The clinical characteristics of participants and patients excluded were described. Chi square test and Student's *t* test were employed to compare the demographic and clinical variables between participants and excluded patients.

The uni-dimensionality and item fit of the HADS-A was assessed with the Rasch model (Wright and Mok, 2000) using the Winsteps software package, Version 3.04 (MSEA Press, Chicago, IL, USA), which implements an unconditional maximum likelihood procedure (Wright and Panchapakesan, 1969). The Rasch model proposes that responses to a set of items can be explained by a person's ability (or anxiety symptoms in this instance) and by the characteristics of the items. Patients' and item scores are used to 'calibrate' items (i.e. characteristics or symptoms) on a logit scale. Items at one end of the scale are 'easier' and items at the other end are more 'difficult.' The difficulty of individual items is determined by the frequency of endorsement. A logit is the natural log-odds of the level of difficulty of a particular item in relation to all other items in the scale (hierarchy). Rasch analysis also constructs a hierarchy of the respondents ordered by their level of anxiety symptoms.

Uni-dimensonality refers to the single underlying construct measured by items that form a scale; for the HADS-A it is the severity of anxiety symptoms. The adequacy of the fit of each item to the Rasch model is assessed by the mean-square residual goodness of the information weighted fit (INFIT) and outlier-sensitive fit (OUTFIT) (Wright and Masters, 1982), which are measures providing information about the responses given to items around the same difficulty endorsement level as the person's ability. An acceptable range for an

Copyright © 2008 John Wiley & Sons, Ltd

DOI: 10.1002/mpr

INFIT/OUTFIT value is between 0.7 and 1.3 (Wright and Linacre, 1994). A 'mute' item is one with an INFIT/OUTFIT value of less than 0.7 because it does not provide information beyond that provided by the rest of the items of the scale. This can occur when there are several items that are similar or highly correlated, or when one item is dependent on another. In contrast, a 'misfit' item is one with an INFIT/OUTFIT value of greater than 1.3 indicating that the item does not define the same construct as do the rest of the items; it is either a poorly constructed or understood item, or it is ambiguously defined. Items with poor fit statistics may be substituted or removed from the instrument.

DIF analysis (Holland and Wainer, 1993) was performed to determine the variability of item responses across subgroups of patients identified by sex, age, education, pO_2, six-minute walking distance and FEV_1%. Briefly, for the response of each person to each item, the standardized residual of the observed score from what was predicted by the model was calculated. Then each person was classified according to one of the class intervals, giving a set of residuals suitable for a two-way analysis of variance (ANOVA) design. Thus, the statistical test used for detecting DIF was an ANOVA of the person-item deviation residuals with person factors (e.g. age) and class intervals (e.g. group along the trait) as factors. In the DIF analysis, there were two sex and age classes each, namely, male and female, and <75 and >/=75 years. In the pO_2, six-minute walking distance and FEV_1% classes, patients were split by the medians of the sample. The three classes in education were <1, 1–6 and >6 years. In general, two types of DIF can be identified: uniform and non-uniform DIF. With the former, there is a constant difference between groups in the probability of affirming an item (or category) an item with higher category across the trait (ANOVA main effect), and with the latter the difference varies across the trait (ANOVA interaction effect). Because multiple tests were performed, the level of significance of 0.05 was adjusted by Bonferroni correction to 0.0071 (0.05/7). The WINSTEPS software, Version 3.04, and RUMM2020 software (Rumm Laboratory Pty Ltd, Perth, Australia) were used to perform the analysis.

Results

Characteristics of the sample

The recruitment of patients took place from November 2003 to December 2004. Of the 251 COPD patients admitted, 166 (66.1%) participated and 85 (33.9%) patients were excluded from the study. The reasons for exclusion were refusal to participate in the study (n = 37, 43.5%), physical frailty (n = 32, 37.6%), a MMSE score less than 15 (n = 10, 11.8%), and inability to give consent (n = 8, 9.4%). Compared to the participating group, the excluded group had a higher proportion of female subjects (48.2% versus 22.9%, $p < 0.001$), was older [80.6 ± 8.7 (standard deviation) versus 75.5 ± 8.5 years, $p < 0.001$], and had a shorter duration of COPD (5.4 ± 4.2 versus 7.4 ± 8.4 years, $p = 0.02$).

Of the participating patients, 128 (77.1%) were male, 106 (66.3%) patients were married and 156 (94.0%) patients were retired. The mean age of the sample ranged from 44 to 95 years; the average duration of education was 3.7 ± 3.9 years. The pO_2, six-minute walking distance, FVC%, and FEV_1% were recorded in 152 (91.6%), 139 (83.7%) and 118 (71.1%) patients. The medians of the pO_2, six-minute walking distance, FVC% and FEV_1% were 10.7 kilo Pascals, 134.5 meters, 56.0% and 37.5%, respectively. The mean HADS-A and MMSE scores of the entire sample were 3.5 ± 4.1 and 23.8 ± 4.3, respectively. Twenty-five (15.1%) out of the 166 patients had a HADS-A score of eight or more.

Psychometrics of the HADS-A

The item-person map of the HADS-A is shown in Figure 1. The mean logit score for the sample was –1.61 ± 0.70 indicating that the seven items (mean logit 0.00 ± 0.16) on average were difficult to be endorsed with higher categories (items 2 and 3) in the sample. Items 2 and 6 were misfit items and item 7 was a mute item (Table 1). The HADS-A scores per personal factors are shown in Table 2. None of the items of the HADS-A had a significant DIF for age, sex, education, six-minute walking distance and FEV_1%. One item ('I get a sudden feeling of panic') displayed uniform DIF for pO_2 (Table 1). Patients who affirmed this item tended to have lower pO_2.

Discussion

Overall, the HADS-A items form a uni-dimensional hierarchy, but one item was mute (item 7) and two items (items 2 and 6) were of borderline misfit. A recently published Rasch analysis also revealed that item 6, but not items 2 or 7, of the HADS-A demonstrated misfit and the removal of this item had little impact on screening (Smith et al., 2006).

Copyright © 2008 John Wiley & Sons, Ltd
DOI: 10.1002/mpr

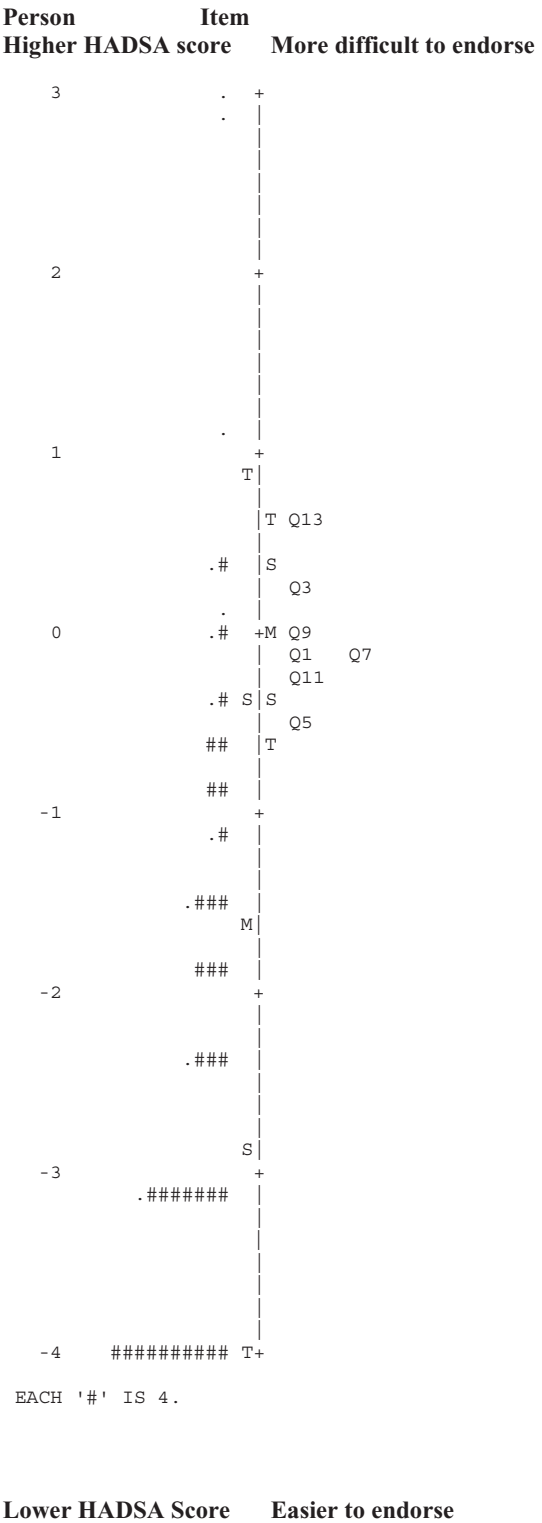

Figure 1. Item-person map of the anxiety subscale of the Hospital Anxiety and Depression Scale (HADS-A).

The finding of modest misfit on a single item of HADS-A should be interpreted as a large number of factor analytic studies have confirmed the factor of HADS-A variance (White et al., 1999; Dagnan et al., 2000). Even though Smith et al. (2002) suggested that misfit items may be substituted or removed to increase the representativeness and measurement precision of a scale. Further evaluation is required before considering the removal of any item of the scale.

For any unbiased item, the probability of a person affirming an item at a given level of anxiety should be the same for different classes of demographic or severity of disease variables. Items that do not yield the same item response function for two or more groups display DIF and are violating the requirement of unidimensionality (Tennant et al., 2004). In this study, only one item of the HADS-A displayed DIF for pO_2 and none of the items had a significant DIF for age, sex, education or other indicators of severity of disease. Osborne et al. (2004) reported that items 2 and 5, but not item 7, exhibited small DIF and would marginally increase the apparent excess anxiety in women with breast cancer compared with population women.

To the best of our knowledge, this is the first DIF analysis of the HADS-A in patients with a chronic lung disease. Our observation that the item on panic feelings had DIF by pO_2 may be explained by the suffocation false alarm model (Klein, 1993). In this model, panic attacks could be triggered by a hypersensitive brainstem autonomic control mechanism that may fire spontaneously or after minor suffocation-related stimulation, thereby initiating both panic and a sensation of dyspnea. Falls in pO_2 can result in feeling of suffocation, and increased ventilation. The only study that tested panic disorder patients' sensitivity to hypoxia demonstrated that they reacted to a 12% oxygen challenge (air normally contains 20%) with more anxiety and ventilatory responses than controls (Beck et al., 1999).

One of the limitations of this study is that only the Chinese version of HADS-A was evaluated and it was not self-rated, the sample consisted of Chinese only, and the sample size is small which may lead to type II error, the attrition rate was quite high and many female and older patients were excluded. Only 15% of our sample had clinically relevant anxiety (HADS-A score of eight or more), which was much lower than the figure of 47% to 50% previously reported in patients with COPD (Dowson et al., 2001; Jones et al., 1989). It would

Copyright © 2008 John Wiley & Sons, Ltd

DOI: 10.1002/mpr

Table 1. The INFIT/OUTFIT statistics and DIF of the anxiety subscale of the Hospital Anxiety and Depression Scale (HADS-A) by pO_2, six minute walking distance and FEV_1%

Item	INFIT Mean square	OUTFIT Mean square	pO2 *n* = 152		Six minute walking distance *n* = 139		FEV_1% *n* = 118	
			U	NU	U	NU	U	NU
1. I feel tense or wound up.	0.84	0.81	0.6397	0.8704	0.8706	0.3290	0.5901	0.8939
2. I get a sort of frightened feeling as if something awful is about to happen.	1.32	1.08	0.2567	0.3677	0.4959	0.2075	0.9054	0.5094
3. Worrying thoughts go through my mind.	0.83	0.88	0.6491	0.7238	0.3063	0.5343	0.5838	0.8347
4. I can sit at ease and feel relaxed.	1.17	1.12	0.9343	0.1678	0.0213	0.2037	0.3265	0.2216
5. I get a sort of frightened feeling like 'butterflies' in the stomach.	0.88	0.82	0.1331	0.3635	0.5857	0.7910	0.7065	0.0527
6. I feel restless as if I have to be on the move.	1.28	1.36	0.3816	0.7460	0.0083	0.7574	0.4976	0.9484
7. I get sudden feelings of panic.	0.82	0.58	0.0004a	0.0382	0.3144	0.5864	0.4910	0.6878

Note: The Bonferroni adjusted significance level for each uniform (U) and non-uniform (NU) DIF. The level of significance for all comparisons was 0.0071.

Table 2. The mean and standard deviation of the total HADS-A score per personal factors

	Male (*n* = 128)		Female (*n* = 38)
	3.4 ± 3.7		3.9 ± 5.3
Age (years)	<75 (*n*= 72)		≥75 (*n* = 94)
	4.1 ± 4.4		3.1 ± 3.8
FEV_1%	<37.5 (*n* = 78)		≥37.5 (*n* = 78)
	4.1 ± 4.3		3.3 ± 3.6
pO_2 (kilo Pascals)	<10.7 (*n* = 78)		≥10.7 (*n* = 78)
	3.3 ± 3.5		3.7 ± 4.1
Six minute walking distance (meters)	<134.5 (*n* = 78)		≥134.5 (*n* = 78)
	3.5 ± 4.0		2.8 ± 3.2
Education (years)	<1 (*n* = 56)	1–6 (*n* = 80)	>6 (*n* = 0)
	3.7 ± 5.0	3.4 ± 3.5	3.5 ± 3.8

be useful to replicate our findings in another population with higher levels of anxiety. In addition, the pO_2 and lung function parameters were missing for a proportion of patients.

In conclusion, this study found that one item of the HADS-A has significant item bias for the severity of disease in patients with COPD.

Copyright © 2008 John Wiley & Sons, Ltd

DOI: 10.1002/mpr

Acknowledgments

This work was supported by a research grant from the Pneumoconiosis Compensation Fund Board of Hong Kong 2003.

Declaration of Interests

The authors have no competing interests.

References

Azocar F, Arean P, Miranda J, Munoz RF. Differential item functioning in a Spanish translation of the Beck Depression Inventory. J Clin Psychol 2001; 57: 355–65. DOI: 10.1002/jclp.1017

Beck JG, Ohtake PJ, Shipherd JC. Exaggerated anxiety is not unique to CO2 in panic disorder: a comparison of hypercapnic and hypoxic challenges. J Abnorm Psychol 1999; 108: 473–82. DOI: 10.1037/0021-843X.108.3.473

Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002; 52: 69–77. DOI: 10.1016/S0022-3999(01)00296-3.

Chan KS, Orlando M, Ghosh-Dastidar B, Duan N, Sherbourne CD. The interview mode effect on the Center for Epidemiological Studies Depression (CES-D) scale: an item response theory analysis. Med Care 2004; 42: 281–9. DOI: 10.1097/01.mlr.0000115632.78486.1f

Chiu HF, Lee CB, Chung D, Kwong PK. Reliability and validity of the Cantonese version of the Mini-Mental State Examination: a preliminary study. HK J Psychiatry 1994; 4: 25–8.

Crane PK, van Belle G, Larson EB. Test bias in a cognitive test: differential item functioning in the CASI. Stat Med 2004; 23: 241–56. DOI: 10.1002/sim.1713.

Dagnan D, Chadwick P, Trower P. Psychometric properties of the Hospital Anxiety and Depression Scale with a population of members of a depression self-help group. Br J Med Psychol 2000; 73(1): 129–37. DOI: 10.1348/000711200160255.

Dowson C, Laing R, Barraclough R, Town I, Mulder R, Norris K, Drennan C. The use of the Hospital Anxiety and Depression Scale (HADS) in patients with chronic obstructive pulmonary disease: a pilot study. NZ Med J 2001; 114: 447–9.

Garuti G, Cilione C, Dell'Orso D, Gorini P, Lorenzi MC, Totaro L, Cirelli G, Clini E. Impact of comprehensive pulmonary rehabilitation on anxiety and depression in hospitalized COPD patients. Monaldi Archives for Chest Disease 2003; 59: 56–61.

Guyatt GH, Sullivan MJ, Thompson PJ, Fallen EL, Pugsley SO, Taylor DW, Berman LB. The six minute walk: a new measure of exercise capacity in patients with chronic heart failure. Can Med Assoc J 1985; 132: 910–1023.

Herrman C. International experiences with the Hospital Anxiety and Depression Scale – a review of validation data and clinical results. J Psychosom Res 1997; 42: 17–41. DOI: 10.1016/S0022-3999(96)00216-4.

Hinz A, Schwarz R. Anxiety and depression in the general population: standardised values of the hospital anxiety and depression scale. Psychotherapie, Psychosomatik, Medizinische Psychologie 2001; 51: 193–200.

Hinz A, Hueber B, Schreinicke G, Seibt R. Temporal stability of psychophysiological response patterns: concepts and statistical tools. Int J Psychophysiol 2002; 44(1): 57–65. DOI: 10.1016/S0167-8760(01)00191-X.

Holland PW, Wainer H (eds). Differential Item Functioning. Hillsdale, NJ: Lawrence Erlbaum, 1993.

Jones PW, Baveystock CM, Littlejohns P. Relationships between general health measured with the sickness impact profile and respiratory symptoms, physiological measures, and mood in patients with chronic airflow limitation. Am Rev Respir Dis 1989; 140: 1538–43.

Jones RN, Gallo JJ. Education and sex differences in the mini-mental state examination: effects of differential item functioning. J Gerontol B Psychol Sci Soc Sci 2002; 57: 548–58.

Klein DF. False suffocation alarms, spontaneous panics, and related conditions. An integrative hypothesis. Arch Gen Psychiatry 1993; 50: 306–17.

Leung CM, Wing YK, Kwong PK, Lo A, Shum K. Validation of the Chinese version of the Hospital Anxiety and Depression Scale and comparison with the Hamilton Rating Scale of Depression. Acta Psychiatr Scand 1999; 100: 456–461.

Oga T, Nishimura K, Tsukino M, Hajiro T, Ikeda A, Mishima M. Relationship between different indices of exercise capacity and clinical measures in patients with chronic obstructive pulmonary disease. Heart and Lung: J Crit Care 2002; 31: 374–81. DOI: 10.1067/mhl.2002.127941.

Osborne RH, Elsworth GR, Sprangers MA, Oort FJ, Hopper JL. The value of the Hospital Anxiety and Depression Scale (HADS) for comparing women with early onset breast cancer with population-based reference women. Qual Life Res 2004; 13: 191–206.

Scheuneman JD. A method of assessing bias in test items. J Educ Measurement 1979; 16: 143–52.

Smith AB, Wright EP, Rush R, Stark DP, Velikova G, Selby PJ. Rasch analysis of the dimensional structure of the Hospital Anxiety and Depression Scale. Psycho-Oncology 2006; 15: 817–27. DOI: 10.1002/pon.1015.

Smith EV Jr, Conrad KM, Chang K, Piazza J. An introduction to Rasch measurement for scale development and person assessment. J Nurs Measurement 2002; 10: 189–206. DOI: 10.1891/jnum.10.3.189.52562.

Smith LL, Reise SP. Gender differences on negative affectivity: an IRT study of differential item functioning on the Multidimensional Personality Questionnaire Stress Reaction Scale. J Pers Soc Psychol 1998; 75: 1350–62. DOI: 10.1037/0022-3514.75.5.1350.

Smoller JW, Pollack MH, Otto MW, Rosenbaum JF, Kradin RL. Panic anxiety, dyspnea, and respiratory disease:

Copyright © 2008 John Wiley & Sons, Ltd

theoretical and clinical considerations. Am J Respir Crit Care Med 1996; 154: 6–17.

Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999; 48: 1507–15. DOI: 10.1016/S0277-9536(99)00045-3.

Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, Lawton G, Simone A, Carter J, Lundgren-Nilsson Å, Tripolski M, Ring H, Biering-Sørensen F, Marincek C, Burger H, Phillips S. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004; 42(Suppl. 1): 137–48. DOI: 10.1097/01.mlr.0000103529.63132.77.

White D, Leach C, Sims R, Atkinson M, Cottrell D. Validation of the Hospital Anxiety and Depression Scale for use with adolescents. Br J Psychiatry 1999; 175: 452–4.

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

Wright BD, Masters GN. Rating Scale Analysis. Chicago, IL: Mesa Press, 1982.

Wright BD, Mok M. Rasch models overview. J Appl Measurement 2000; 1: 83–106.

Wright BD, Panchapakesan N. A procedure for sample-free item analysis. Educ Psychol Measurement 1969; 29: 23–48. DOI: 10.1177/001316446902900102.

Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983; 67: 361–70.

Correspondence: W. K. Tang, Department of Psychiatry, Shatin Hospital, Shatin, N.T., Hong Kong SAR, China.
Telephone +852-2636-7760
Fax +852-2648-3394
Email: tangwk@cuhk.edu.hk

Copyright © 2008 John Wiley & Sons, Ltd