The Reliability and Validity of the “Activity and Participation” Component in the Brief ICF Core Set for Chronic Obstructive Pulmonary Diseases Based on Rasch Analysis

Purpose: To analyze the reliability and validity of the “activity and participation” component of the brief international classification of functioning, disability and health (ICF) core set for chronic obstructive pulmonary disease (COPD) using a Multi-faceted Rasch model.

Patients and Methods: A total of 103 patients with COPD were selected by two raters to evaluate their ability levels in the four categories of the “activity and participation” component of brief ICF core set for COPD. The Multi-faceted Rasch model was used for data analysis. The analysis software used FACETS (Minifac) 3.67.0.

Results: The “activity and participation” of brief ICF core set for COPD had a high internal consistency (separation index of 5.08, reliability of 0.96, P <0.05) and good inter-rater reliability (mean-square fit statistic range was 0.97–1.04, the separation index was 0.00, the reliability was 0.00, P >0.05), the construct validity was good (mean-square fit statistic range was 0.79–1.36), and the consistency of each category measurement was high (the separation index was 1.70, the reliability was 0.74).

Conclusion: The “activity and participation” of brief ICF core set for COPD has good reliability and validity, which can be used to test the daily activities of patients with COPD.

Keywords: the international classification of functioning, disability and health, chronic obstructive pulmonary disease, reliability, validity, Multi-faceted Rasch model

Introduction

The International Classification of Functioning, Disability and Health (ICF) was issued by the World Health Organization (WHO) in 2001 and approved by the United Nations for international use as a further development of the International Classification of Injuries, Disabilities and Disabilities (ICIDH, 1980). It was a classification of health and health-related fields. The ICF was based on a comprehensive biopsychosocial function model and for the first time achieved standardized descriptions of function and disability, taking into account both personal and environmental factors (background factors). As a result, a paradigm shift has occurred from the classification of “illness consequences” to the classification of “healthy ingredients”. Different from the causal linear relationship between the factors reflected in the previous International Classification of Disease (ICD), the ICF emphasized the dynamic and interactive relationship between the various constituent factors that were interrelated and affect each other. The ultimate
The brief and comprehensive ICF core sets for Dove ≥ and Wright component of the After ICF training and based on the de Therefore, the analysis of the metrological char- both considered 100 as the component was the most characteristic one in pay attention to the overall status of patients, Therefore, it was appropriate to use the MFR model for the psychometric measurement of the brief ICF core set for COPD.

Since ICF was mainly used for disability and rehabilitation management, its main purpose was to help patients overcome the limitations of activities and participation. 13,14 “Activity and participation” component was the most characteristic one in brief ICF core set for COPD. 6,14 Therefore, this study will focus on the “activity and participation” component of the brief ICF core set for COPD. The MFR will be used to analyze its reliability and validity to deeply discover its psychometric characteristic and provide a reference to the assessment and application of other ICF core set.

Patients and Methods
Study Design
The study was designed as a cross-sectional survey. The research site was Zhongshan Hospital, Fudan University. The study has been approved by the Ethics Committee of Zhongshan Hospital, Fudan University with approval number B2018-216R. The study has been registered with the Chinese Clinical Trial Registry under the registration number ChiCTR1900021503.

Subjects
The study subjects were patients with COPD who visi- ted the Zhongshan Hospital, Fudan University from April to October 2019. The inclusion criteria were: (1) meeting the diagnostic criteria for COPD, and the diag- nostic criteria refer to the Global Initiative for Chronic Obstructive Lung Disease (GOLD); (2) age ≥18 years; (3) signing informed consent. Exclusion criteria were: (1) cognitive dysfunction; (2) complicated with severe primary diseases such as cardiovascular, liver, kidney, and brain.

Linacre 15 and Wright 16 both considered 100 as the minimum sample size suitable for Rasch analysis. Therefore, the target minimum sample size for this study was 100.

Study Tools
Demographic Characteristics Questionnaire
It included general information such as the patient’s gen- der, age, education level, marital status, income, occupation, and place of residence, as well as disease information such as time of illness, age of illness, smoking history, and severity of disease.
The Brief ICF Core Set for COPD: “Activity and Participation”
The four categories in the “activity and participation” included d230: carrying out daily routine, d450: walking, d455: moving around and d640: doing housework. ICF’s category qualifiers were rated as 5 levels, 0, 1, 2, 3, and 4, respectively, indicated no difficulty, mild difficulty, moderate difficulty, severe difficulty, and complete difficulty. In addition, eight and nine indicated unspecified and not applicable, respectively. Table 1 shows specific qualifiers assignments.

Data Collection
In the previous research, researchers have worked out detailed evaluation rules based on the ICF qualifiers, using Delphi method, quantified data collection and other methods, and trained the raters to master the ICF evaluation rules and precautions.

Data collection and evaluation were performed by two raters with relevant research experience. After confirming that the patients met the inclusion criteria, raters explained the purpose and content of the study to them, and helped patients sign the informed consent. They used the form of face-to-face interviews to assess the same patient, respectively, in accordance with the assessment rules and refer to the results of the patients’ examinations. The time interval between the completion of the assessment by two raters shall not exceed 3 days.

Data Analysis
The data were analyzed using IBM SPSS Statistics version 21.0. The baseline data were presented as mean ± standard deviation [SD] or proportion (%). MFR was used for data analysis. The data were divided into three aspects: rater severity, and assessed patient ability, and the difficulty of the four categories of the brief ICF core set “activity and participation” for COPD. Multi-faceted Rasch analysis can consider multiple factors affecting the measurement simultaneously. The analysis software used FACETS (Minifac) 3.67.0.

Internal Consistency Reliability
Internal consistency reliability was expressed by the separation statistic and reliability index. The separation index indicated the discriminating ability of the tool. A separation index greater than 5 indicated good discrimination. The reliability index was equivalent to Cronbach’s α, which was used to test internal consistency, and greater than 0.91 indicated that research tools had good internal consistency.

Inter-Rater Reliability
The criterion for rater consistency was whether their distribution of scores for patients with COPD conformed to the MFR. The characteristics of the data with a higher degree of fit to MFR were: (1) the scores for the more difficult items were lower, and the scores for the less difficult items were higher. (2) Patients with higher levels scored higher, patients with lower levels scored lower.

The evaluation index was: (1) information-weighted mean-square fit statistic (Infit MnSq), conventional mean-square fit statistic (Outfit MnSq), and standardized score residuals (Z). MnSq was in the 0.5–1.5 range, and Z < 2 was considered to be a good fit. (2) Measure and Model S.E., Measure was used to check the severity of raters. The acceptable range of measure was ±0.5 logits. S.E.<0.30. (3) The separation index was used to indicate whether the difference between raters was greater than the measurement error. A separation index of <2 indicated that there was no significant difference and raters were consistent.

Validity
In order for a measurement tool to be useful, it must be relevant and consistent with what is being measured. Fit was used to monitor how relevant the tool was to its use.

| Assignments | Meaning Assignments | Degree of Obstacle |
|-------------|---------------------|-------------------|
| 0           | No problem (none, lack, trivial . . .) | 0–4% |
| 1           | Mild problem (slightly, very low . . .) | 5–24% |
| 2           | Moderate problems (moderate, average . . .) | 25–49% |
| 3           | Serious problems (high, very . . .) | 50–95% |
| 4           | Total problems (all . . .) | 96–100% |
| 8           | Unspecified | The information currently available does not adequately determine whether the patient has a problem or the severity |
| 9           | Not applicable | This category does not apply to patients |

Notes: The percentages indicate the extent of the barriers to activities and participation.
Low-fitness statistics indicated that a tool may not provide useful information for measurement, while high-fitness statistics indicated that the tool was unclear or was being used in an unpredictable way.\textsuperscript{21} Construct validity was tested by fitting analysis. Mean Square (MnSq) range of 0.5–1.5 was considered to be a good fit.\textsuperscript{21} Out-of-range fitting data indicated that the measurement item may not be measured scientifically, ensuring measurement quality, and items with a high degree of fit will reduce the overall measurement effect.\textsuperscript{21} Separation index and reliability index were used to comprehensively represent the construct validity of the measurement tool.\textsuperscript{21}

\section*{Results}

\subsection*{Demographic Characteristics}

A total of 103 patients with COPD were included in this study, including 93 males and 10 females. The average age was 65.99 ± 8.93 years, and the average disease duration was 4.53 ± 5.67 years. The general information and disease information of 103 patients are shown in Table 2.

\subsection*{Reliability}

\subsubsection*{Internal Consistency reliability}

The distribution of the item difficulty, rater severity, and patient level of the four categories in the “activity and participation” component are shown in Figure 1. The internal consistency reliability test results are shown in Table 3. The Infit MnSq and Outfit MnSq were both 0.98, between 0.5 and 1.5, the Z value was <2. It showed that it fitted well with the MFR. The separation index was 5.08, >5, the reliability was 0.96, >0.9, and P value was 0.00, <0.05. It showed that this component had good discrimination, and it also can distinguish the activity level of different patients.

\subsubsection*{Inter-Rater Reliability}

The results of the inter-rater reliability test are shown in Table 4. The measures between two raters were −0.03 logits and 0.03 logits, with S.E. of 0.10, <0.3, which were within the allowable range. Infit MnSq and Outfit MnSq were both at 0.5–1.5, Z was <2, indicating that the severity of the rater fitted well with MFR. The separation index was 0.00 and <2, which meant that there was no significant difference between raters. In addition, the P value was 0.68, which further indicated that the difference between raters was not significant, and the scores between raters were relatively consistent.

\subsection*{Validity}

Construct validity was tested by fitting analysis. The results are shown in Table 5. From the results, it can be seen that the fit of the four items was good. And the separation index was 1.7, <2, which indicated that the four items had a higher consistency in measuring patients’ activity.

\section*{Discussion}

The brief ICF core set for COPD was an evaluation tool developed by WHO through a series of expert consultations, expert consensus meetings, and literature reviews to assess the overall function and activity level of patients with COPD.\textsuperscript{5,6} It had a high degree of authority and credibility.
The study did not repeat the content validity test, and analyzed its internal consistency reliability, inter-rater reliability, and construct validity. The analysis results showed that all aspects of item, rater, patient fitted MFR better. It can be seen from the facet map of the component "activity and participation" that the distribution of raters and items were relatively concentrated, which indicated that the assessment standards of the raters and the internal standards of the items were relatively uniform. It has been preliminary proved that the “activity and participation” components of brief ICF core set for COPD have good reliability and validity.

The theoretical basis of Rasch model was Item Response Theory, which meant the ability of the subject was a potential trait of the subject, and it had nothing to do

Figure 1 Facet map of the component “activity and participation”.
Notes: g: rater 1; l: rater 2; d230: carrying out daily routine; d450: walking; d455: moving around; d640: doing housework; *: the patient with COPD.
The model can estimate subject ability, item difficulty, and the severity of rater at the same time. At the same time, it was an ideal theoretical model for measuring ordered questionnaires. MFR can distinguish and isolate the interaction of each facet when measuring each facet, which can greatly reduce the influence of the rater’s subjectivity and the patient’s own ability on various categories. Therefore, using the Rasch model for the reliability and validity tests of the ICF core set was helpful to provide more detailed information about ICF from various aspects such as raters, subjects, and categories. The Rasch model has been widely used in many fields. Some scholars have used the Rasch model on other scales and in the exams. They mainly measure the reliability of inter-rater to understand the effect of differences between raters on overall measurement results. In addition, many scholars have researched ICF-related core sets from various perspectives using the Rasch model to guide the development of ICF and its clinical application. In this study, the MFR was applied to the “activity and participation” component of brief ICF core set for COPD, in order to find the clinical significance and improvement directions of the brief ICF core set for COPD by measuring its reliability and validity.

The evaluation results of categories in ICF and various core sets were expressed by ICF qualifiers. Although there were qualifiers evaluation standards, indicators that were used for evaluation of each category and the detailed classification criteria of qualifiers had no clear evaluation rules as a guide. Therefore, the evaluation results were greatly affected by factors such as the raters’ understanding of various categories, the basis for dividing the qualifiers, and the professional background of raters. In order to avoid the influence of these bias factors, unified category evaluation rules have been developed in the early stage of the study to guide the evaluation. Before the start of the study, two raters have been uniformly trained to ensure their proficiency in using the evaluation rules and fully ensure the evaluation of standardization and unification. The research results also showed that the separation index between the raters was less than 2. The severity of the two raters was more consistent, and there was no obvious difference. It can also be seen from the facet map of the component “activity and participation” that the two raters, g and l, were concentratedly distributed around the measured value 0. The measurement value 0 was equivalent to the mean of the statistically normal distribution. A higher value above 0 indicated a higher severity of the rater, and a lower value below 0 indicated a lower severity of the rater. It indicated that the two raters had the same severity, and there was no case where the assessment was too severe or too loose.

This research first consulted with some statistical experts to form a model for the development of ICF category evaluation rules. According to the model and

| Measure | Model S.E. | Infit | Z | Outfit |
|---------|-----------|-------|---|--------|
|         |           | MnSq  | Z | MnSq   | Z |
| 0       | 0.71      | 0.98  | -0.4 | 0.98  | -0.4 |
| Separation 5.08 | Strata 7.10 | Reliability 0.96 | $\chi^2=2695.9$ | P=0.00 |

**Abbreviations:** Infit MnSq, information-weighted mean-square fit statistic; Outfit MnSq, conventional mean-square fit statistic; Z, standardized score residuals.

| Rater | Measure | Model S.E. | Infit | Z | Outfit |
|-------|---------|-----------|-------|---|--------|
| g     | -0.03   | 0.10      | 1.04  | 0.5 | 0.99   |
| l     | 0.03    | 0.10      | 1.01  | 0.1 | 0.97   |
| Separation 0.00 | Strata 0.33 | Reliability 0.00 | $\chi^2=0.2$ | P=0.68 |

**Notes:** g: rater 1; l: rater 2.
**Abbreviations:** Infit MnSq, information-weighted mean-square fit statistic; Outfit MnSq, conventional mean-square fit statistic; Z, standardized score residuals.

| Item | Measure | Model S.E. | Infit | Z | Outfit |
|------|---------|-----------|-------|---|--------|
| d455 | 0.24    | 0.14      | 0.83  | -1.7 | 0.79   |
| d230 | 0.20    | 0.14      | 1.36  | 3.3  | 1.29   |
| d450 | -0.11   | 0.14      | 1.06  | 0.6  | 1.01   |
| d640 | -0.34   | 0.14      | 0.84  | -1.7 | 0.83   |
| Separation | Strata | Reliability | $\chi^2=11.7$ | P=0.01 |
| 1.70 | 2.60    | 0.74      |       |      |        |

**Notes:** d455: moving around; d230: carrying out daily routine; d450: walking; d640: doing housework.
**Abbreviations:** Infit MnSq, information-weighted mean-square fit statistic; Outfit MnSq, conventional mean-square fit statistic; Z, standardized score residuals.

The research first consulted with some statistical experts to form a model for the development of ICF category evaluation rules. According to the model and

Table 4 Inter-Rater Reliability Test Results

| Rater | Measure | Model S.E. | Infit | Z | Outfit |
|-------|---------|-----------|-------|---|--------|
| g     | -0.03   | 0.10      | 1.04  | 0.5 | 0.99   |
| l     | 0.03    | 0.10      | 1.01  | 0.1 | 0.97   |
| Separation 0.00 | Strata 0.33 | Reliability 0.00 | $\chi^2=0.2$ | P=0.68 |

**Notes:** g: rater 1; l: rater 2.
**Abbreviations:** Infit MnSq, information-weighted mean-square fit statistic; Outfit MnSq, conventional mean-square fit statistic; Z, standardized score residuals.
the characteristics of each category of the four categories of “activity and participation” components, the evaluation rules for each category of the “activity and participation” component were formed by using the methods of data quantification and classification. The results of the study showed that the data used in these assessment rules were in line with the characteristics of Rasch model, and the inter-rater reliability was good. It proved that the evaluation rules were relatively standard and suitable for guiding clinical staff to use a brief ICF core set for COPD to evaluate patients.

The “activity and participation” component of the brief ICF core set for COPD included four categories: carrying out daily routine, walking, moving around and doing housework. The activities measured in each category were different. From the facet map of the component “activity and participation” in Figure 1, the four categories were concentrically distributed above and below the measured value 0, indicating that the four categories were equally difficult and not too high or too low. In the validity test, Infit MnSq and Outfit MnSq were all between 0.5 and 1.5, which indicated that the fit of this aspect to the MFR model was good. The separation index was 1.7 and <2, indicating that the consistency of the four categories in measuring the activity of patients with COPD was high. And based on the results of internal consistency tests, the component was more distinguishable in measuring the mobility of different patients. It meant the component clearly distinguished the mobility of different patients. At the same time, it was very accurate and appropriate to measure the activity and participation ability of patients with COPD.

This study has the following limitations: (1) Only the “activity and participation” component in the brief ICF core set for COPD was tested for reliability and validity, and the remaining three components were not studied. The remaining three components were body structure, body function, and environmental factors. The measured target abilities differed greatly from each other. Because “activity and participation” was the core component of brief ICF core set for COPD, in this study, we only test the activity ability of brief ICF core set for COPD. Further study can measure the reliability and validity of all the components of the core set to study the overall fit of the core set and the discrimination among the components. (2) Due to the limitation of the study time, only 103 patients with COPD were included in the study, and the sample size can continue to be expanded in the future.

Conclusion
The MFR model was used to test the reliability and validity of the “activity and participation” component of brief ICF core set for COPD. It has been preliminary proved that the component has good reliability and validity, and can be used to measure their daily activities in patients with COPD.

Data Sharing Statement
This trial was conducted in accordance with the Declaration of Helsinki. No further data will be shared.

Author Contributions
All authors contributed to the acquisition, analysis and interpretation of data, drafting and revising the article critically for important intellectual content, gave final approval of the version and agreed to be accountable for all aspects of the work.

Funding
This work was supported by grants from the National Key Research and Development Program of China (No. 2018YFC1313600) and the National Natural Science Foundation of China (Nos. 81570081 and 81770083).

Disclosure
The authors report no conflicts of interest in this work.

References
1. Schwarzkopf SR, Grill E, Dreinhöfer KE. Introduction to the international classification of functioning, disability and health (ICF). Unfallchirurg. 2010;113(6):436–440. doi:10.1007/s00113-010-1742-7
2. Saleebey PW. An introduction to the international classification of functioning, disability and health (ICF). Int J Disabil Hum Dev. 2016;15(1):1–3. doi:10.1515/ijdhd-2015-0027
3. WHO. International classification of functioning, disability and health (ICF). Available from: https://www.who.int/classifications/icf/en/. Accessed May 7, 2020.
4. Dahl TH. International classification of functioning, disability and health: an introduction and discussion of its potential impact on rehabilitation services and research. J Rehabil Med. 2002;34(5):201–204. doi:10.1080/165019702760279170
5. Cieza A, Ewert T, Ustün TB, Chatterji S, Kostanjsek N, Stucki G. Development of ICF core sets for patients with chronic conditions. J Rehabil Med. 2004;36(44 Suppl):9–11. doi:10.1080/16501960410016046
6. Gerdol S, Stoll T, Cieza A. ICF core sets for obstructive pulmonary diseases. J Rehabil Med. 2004;36(44 Suppl):114–120. doi:10.1080/16501960410016794
7. Weigl M, Cieza A, Andersen C, Kollerits B, Amann E, Stucki G. Identification of relevant ICF categories in patients with chronic health conditions: a delphi exercise. J Rehabil Med. 2004;36(44 Suppl):12–21. doi:10.1080/16501960410015443
8. Wolff B, Cieza A, Parentin A, et al. Identifying the concepts contained in outcome measures of clinical trials on four internal disorders using the international classification of functioning, disability and health as a reference. *J Rehabil Med.* 2004;36(44 Suppl):37–42. doi:10.1080/16501960410015407

9. Ewert T, Fuesl M, Cieza A, et al. Identification of the most common patient problems in patients with chronic conditions using the ICF checklist. *J Rehabil Med.* 2004;36(44 Suppl):22–29. doi:10.1080/16501960410015362

10. Ballert CS, Stucki G, Biering-Sorensen F, Cieza A. Towards the development of clinical measures for spinal cord injury based on the international classification of functioning, disability and health with rasch analyses. *Arch Phys Med Rehabil.* 2014;95(9):1685–1694. doi:10.1016/j.apmr.2014.05.006

11. Tesio L. Measuring behaviours and perceptions: rasch analysis as a tool for rehabilitation research. *J Rehabil Med.* 2003;35(3):105–115. doi:10.1080/16501970310010448

12. Lautamo T. Heikkila MJSJoOT. Inter-rater reliability of the play assessment for group settings. *Scand J Occup Ther.* 2011;18(1):3–10.

13. Stucki G, Ewert T, Cieza A. Value and application of the ICF in rehabilitation medicine. *Disabil Rehabil.* 2002;24(17):932–938. doi:10.1080/09638280210148594

14. Marques A, Jacome C, Gabriel R, Figueiredo D. Comprehensive ICF core set for obstructive pulmonary diseases: validation of the activities and participation component through the patient’s perspective. *Disabil Rehabil.* 2013;35(20):1686–1691. doi:10.3109/09638288.2012.750691

15. Linacre JM. Sample size and item calibration stability. *Rasch Meas Trans.* 1994;7(3):328.

16. Wright BD. Misunderstanding the rasch model. *J Educ Meas.* 1977;14(3):219–225. doi:10.1111/j.1745-3984.1977.tb00039.x

17. Fisher WP. Rating scale instrument quality criteria. *Rasch Meas Trans.* 2007;27.

18. Lindner HYN, Linacre JM, Norling Hermansson LM. Assessment of capacity for myoelectric control: evaluation of construct and rating scale. *J Rehabil Med.* 2009;41(6):467–474. doi:10.2340/16501977-0361

19. Linacre JM. Sample size and item calibration [or person measure] stability. *Rasch Measurement Transactions;* 1994. Available from: https://www.rasch.org/rmt/rmt74m.htm. Accessed May 7, 2020.

20. Wright B. Reliability and separation. *Rasch Measurement Transactions;* 1996. Available from: https://www.rasch.org/rmt/rmt94n.htm. Accessed May 7, 2020.

21. Scanlan JN, Lannin NA, Hoffmann T, Stanley M, McDonald R. Impact of adjusting for inter-rater variability in conference abstract ranking and selection processes. *Aust Occup Ther J.* 2018;65(1):54–62. doi:10.1111/1440-1630.12440

22. Linacre JM. New approaches to determining reliability and validity. *Res Q Exerc Sport.* 2000;71(2):129–136. doi:10.1080/02701367.2000.11082796

23. Li K, Yan T, You L, et al. Psychometric properties of the international classification of functioning, disability and health set for spinal cord injury based on Rasch analysis. *Disabil Rehabil.* 2018;40(3):338–345. doi:10.1080/09638288.2016.1250169

24. Tavakol M, Pinner G. Using the Many-Facet Rasch Model to analyse and evaluate the quality of objective structured clinical examination: a non-experimental cross-sectional design. *BMJ Open.* 2019;9(9):e029208. doi:10.1136/bmjopen-2019-029208

25. Tseng W-T, Su T-Y, Nix J-ML. Validating translation test items via the Many-Facet Rasch Model. *Psychol Rep.* 2019;122(2):748–772.

26. Bagraith KS, Strong J, Meredith PJ, McPhail SM. Rasch analysis supported the construct validity of self-report measures of activity and participation derived from patient ratings of the ICF low back pain core set. *J Clin Epidemiol.* 2017;84:161–172. doi:10.1016/j.jclinepi.2016.12.016

27. Prodigier B, Stucki G, Coenen M, Tennial A. The measurement of functioning using the international classification of functioning, disability and health: comparing qualifier ratings with existing health status instruments. *Disabil Rehabil.* 2019;41(5):541–548. doi:10.1080/09638288.2017.1381186

28. Maini M, Nocentini U, Prevedini A, Giardini A, Muscolo E. An Italian experience in the ICF implementation in rehabilitation: preliminary theoretical and practical considerations. *Disabil Rehabil.* 2008;30(15):1146–1152. doi:10.1080/09638280701478397

29. Lautamo T, Heikkila M. Inter-rater reliability of the play assessment for group settings. *Scand J Occup Ther.* 2011;18(1):3–10. doi:10.3109/110381209034800048