

ORIGINAL ARTICLE

Cross-cultural validation of the work functioning impairment scale (WFun) among Japanese, English, and Chinese versions using Rasch analysis

Yoshihisa Fujino¹  | Ning Liu²  | Odgerel Chimed-Ochir¹  | Makoto Okawara¹ | Tomohiro Ishimaru³  | Tatsuhiko Kubo¹ 

¹Department of Environmental Epidemiology, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, Kitakyushu, Japan

²Department of Preventive Medicine and Community Health, University of Occupational and Environmental Health, Kitakyushu, Japan

³Department of Occupational Health Practice and Management, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, Kitakyushu, Japan

Correspondence

Yoshihisa Fujino, Department of Environmental Epidemiology, Institute of Industrial Ecological Sciences, University of Occupational and Environmental Health, Japan 1-1, Iseigaoka, Yahatanishiku, Kitakyushu, 807-8555, Japan.
Email: zenq@med.uoeh-u.ac.jp

Funding information

UOEH Research Grant for Promotion of Occupational Health

Abstract

Objectives: The work functioning impairment scale (WFun) was developed to measure the degree of work functioning impairment in Japanese workers based on the Rasch model. Given that the number of foreign workers employed in Japan and abroad has increased in recent years, a multilingual questionnaire is becoming increasingly necessary to investigate work functioning impairment in these workers. The purpose of this study was to verify the cross-cultural validity of WFun between Japanese, Chinese, and English versions.

Methods: A cross-sectional study was conducted in two stages. First, the Chinese and English versions of WFun were created. Second, an internet survey was conducted among 1000 Japanese, 400 Chinese, and 300 Americans. Estimates and standard errors of an individual's ability and item difficulty were calculated using the Rasch model. Differential item functioning (DIF) and differential test functioning (DTF) were also examined using Rasch model analyses.

Results: The effect size of DIF for one item in the English version exceeded 0.5 logit, indicating the presence of some DIF. In contrast, the effect sizes of DIF for all other items were below 0.5 logit, indicating that the influence of DIF was negligible. Furthermore, Rasch measurements according to the raw score for each version of WFun showed strong agreement among the three versions, with an intraclass correlation of 0.98 (95% confidence interval: 0.97-0.99), indicating the absence of DTF.

Conclusions: Our findings indicate that the English, Chinese, and Japanese versions of WFun have good comparability.

KEYWORDS

China, Japan, patient outcome assessment, presenteeism, translations, work capacity evaluation

1 | INTRODUCTION

Interest in presenteeism is increasing. Presenteeism refers to the practice of continuing to work despite being sick or

being of poor health.^{1,2} Reports have shown that presenteeism is significantly associated with productivity loss.³⁻¹⁰ Consequently, many have sought to measure the magnitude of the impact of presenteeism on worker productivity. These

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2019 The Authors. *Journal of Occupational Health* published by John Wiley & Sons Australia, Ltd on behalf of The Japan Society for Occupational Health

efforts have led to the development of evaluation tools such as the Stanford Presenteeism Scale, Work Productivity and Activity Impairment Questionnaire, Work Limitation Questionnaire, and Work Performance Questionnaire.¹¹⁻¹⁴

The work functioning impairment scale (WFun) was developed to measure the degree of work functioning impairment in Japanese workers based on the Rasch model.¹⁵⁻¹⁷ That is, WFun endeavors to express the level of worker health problems in terms of the extent to which a worker experiences reduced functioning at work as a consequence of these problems. In contrast, other presenteeism indexes evaluate productivity based on a number of different factors, such as time not spent on a job, standard of work, amount of work, and personal factors.⁶ WFun has been validated according to Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN).^{18,19} COSMIN establishes and provides recommendations for examining various categories of validity and reliability for measurement instruments.

Cross-cultural validity refers to the degree that the performance of the items in a translated or culturally adjusted patient-reported outcome tool suitably reproduces the performance of items in the source tool. Given that the number of foreign workers employed in Japan and abroad has increased in recent years, there is an increasing need for a multilingual questionnaire to investigate the work functioning impairment in these workers. Furthermore, for future international comparisons, it is necessary to evaluate the validity among foreign workers in general, not just those working in Japanese companies. However, WFun has not been cross-culturally validated.

Questionnaire items may not possess the same function across different cultural groups, and such items are said to show cross-cultural bias or differential item functioning (DIF) according to culture.²⁰⁻²² DIF not only arises from problems with translation but also from cultural heterogeneity. The existence of DIF biases comparability between the same questionnaires written in different languages.

The purpose of this study was to verify the cross-cultural validity of WFun between Japanese, Chinese, and English versions.

2 | METHODS

This cross-sectional study was performed in two stages. First, Chinese and English versions of WFun were created. Second, Rasch analysis and DIF verification were conducted.

2.1 | Cross-cultural adaptation

WFun consists of the following seven items: "I haven't been able to behave socially", "I haven't been able to maintain the quality of my work", "I have had trouble thinking clearly", "I

have taken more rests during my work", "I have felt that my work isn't going well", "I haven't been able to make rational decisions", and "I haven't been proactive about my work". Respondents are required to choose from one of the following five response categories for each item: 1, "not at all"; 2, "one or more days a month"; 3, "about one day a week"; 4, "two or more days a week"; and 5, "almost every day." The final WFun score was the sum of the scores of the 7 items. Scores could range from 7 to 35, with higher scores indicating worse work ability.

Translation was performed based on previously described methods.^{19,23} First, the Japanese version of WFun was translated into the target language by two translators independently. The authors then unified the two resulting translations with consultation. Subsequently, the unified translation was back-translated into Japanese. The final translated versions were then developed according to the contents of the back translation at an expert meeting. This process was performed for both the Chinese and English versions. The Japanese, English, and Chinese versions of WFun are available from the corresponding author upon request.

2.2 | Cross-cultural validation

2.2.1 | Subjects

The Japanese, Chinese, and English versions of WFun were examined using internet surveys. For the Japanese version, the internet survey was conducted in our previous study for the original development of WFun¹⁵; the data were reanalyzed for the purposes of the present study.

The original Japanese version of WFun was examined using an internet survey that targeted 1000 registered Japanese monitors, as described previously.¹⁵ Briefly, we enlisted a commercial testing company to perform an internet test user study. An email was sent to approximately 20 000 of the 2 million registered internet test users asking for participation in the survey. Potential participants were screened for the inclusion of sentences such as "I am currently employed" and "I have some health issues". We excluded workers who did not have any health issues because WFun aims to measure the degree of work functioning impairment due to health problems. Registered users who satisfied these criteria were categorized into five age groups (20s, 30s, 40s, 50s, and 60s) according to sex, with each group containing 100 respondents. The first 1,000 responses were included to attain the target population for each group. Respondents were asked about their age, sex, occupation, and employment type, and provided responses to the WFun items.

Likewise, the Chinese version was examined using an internet survey targeting 400 Chinese respondents aged 20–59 years living in mainland China, and the English version was examined using an internet survey targeting 300

Americans aged 20–59 years living in the United States. American subjects registered to Dynata™, to which approximately 7 million subjects are registered, and Chinese subjects registered to iPanel Online Market Research Ltd., to which approximately 100 000 subjects are registered, were used for the internet survey.

Given that this research uses data that do not include personal information from internet monitors, the need for ethical approval from an ethical committee was waived.

2.2.2 | Statistical analysis

The Rasch model is a common statistical method used to estimate latent ability based on item responses.^{24–26} The Rasch model is a mathematical framework that provides approximations and standard errors of person ability and item difficulty, which are determined on a common equal-interval logit scale. In the Rasch model, one variable is used to approximate person ability (total correct responses by the individual) and item difficulty (total correct responses to an item) to calculate the likelihood of the individual being successful at the item. The Rasch analysis was conducted in WINSTEPS version 4.2.0. Data were fitted to the Rasch rating scale model with the joint maximum likelihood estimation, where all items had equivalent rating scale structures. The magnitude of DIF is known as the DIF contrast or effect size, and indicates logit differences in Rasch model difficulty estimates.²¹ The recommended effect size is typically 0.40–0.60 logit,^{27–29} although standards for important effect size are lacking. In practice, <0.50 logit is used to indicate no DIF as “measures based on item calibration with random deviations up to 0.50 logit are ‘for all practical purposes free from bias.’”^{27–29}

We also evaluated the potential presence of differential test functioning (DTF). Given that a subject is evaluated based on the results from the entire test, it is important to verify whether the existence of DIF affects the evaluation of the whole test.^{29–31} To do this, we approximated Rasch measurements correlating to the raw scores from each of the Japanese, Chinese, and English versions of WFun, and subsequently calculated the absolute consistency using intraclass correlation (ICC) (2,1).³²

3 | RESULTS

Table 1 shows the demographic characteristics of the study subjects. Due to planned sampling, there were no substantial differences in the age or gender of respondents among the three versions of WFun. The proportion of occupation was similar between Japanese and American subjects. Approximately, 50% of the subjects were desk workers. On the other hand, the percentage of desk work was high among the Chinese subjects, with 76% being desk work.

Table 2 shows the estimated Rasch measurements for all groups combined (Japanese, Chinese, Americans) and for each group separately. For Japanese respondents, item 6 (“I haven’t been able to make rational decisions”) had the highest value of 0.40 logit. This indicates that respondents who answered “yes” to item 6 may experience the most severe work functioning impairment. In contrast, item 3 (“I have had trouble thinking clearly”) had the lowest value of –0.38 logit. Similar results were obtained for American respondents. For Chinese respondents, item 1 (“I haven’t been able to behave socially”) had the highest value of 0.55 logit, while item 2 (“I haven’t been able to maintain the quality of my work”) had the lowest value of –0.53 logit.

Table 3 shows the effect size of DIF for the three versions of WFun. Only the effect size for item 4 (“I have taken more rests during my work”) in the English version exceeded 0.5 logit, indicating the presence of some DIF for this item. The effect sizes for all other items were less than 0.5 logit, indicating that DIF was negligible.

Table 4 shows the Rasch measurements according to the raw score from each version of WFun used to determine the potential presence of DTF. The ICC was 0.98 (95% confidence interval: 0.97–0.99), indicating strong agreement among the three versions and, therefore, the absence of DTF.

4 | DISCUSSION

This study examined the cross-cultural validity of Japanese, Chinese, and English versions of WFun. DIF was identified for only one item in the English version, but was negligible for the other items and all items in the Chinese version. Furthermore, there was no DTF among the three versions.

Two processes are necessary when adopting patient-reported outcome measures in multiple languages: cross-cultural adaptation and cross-cultural validation.³³ Cross-cultural

TABLE 1 Basic characteristics of study subjects

	Japanese	Chinese	American
Number of subjects	1000	400	300
Men (%)	50 ^a	50 ^a	50 ^a
Age (mean and SD)	44 (13) ^a	39 (11) ^a	43 (13) ^a
Job type (%)			
Mainly desk work	51	76	55
Mainly work involving interpersonal communication	23	15	22
Mainly physical work	25	9	23

^aEqual number of subjects were assigned to 10-year age groups (20s, 30s, 40s, 50s and 60s) by sex.

TABLE 2 Rasch measurements by different language versions

Item ^a	Total (n = 1700)		Japanese (n = 1000)		Chinese (n = 400)		American (n = 300)	
	item difficulty	SE	item difficulty	SE	item difficulty	SE	item difficulty	SE
q1	0.00	0.03	-0.29	0.05	0.55	0.08	0.22	0.07
q2	-0.24	0.03	-0.20	0.05	-0.53	0.07	-0.09	0.07
q3	-0.30	0.03	-0.38	0.05	-0.15	0.07	-0.31	0.07
q4	0.17	0.04	0.39	0.05	0.16	0.08	-0.29	0.07
q5	-0.19	0.03	-0.15	0.05	-0.47	0.07	-0.03	0.07
q6	0.37	0.04	0.40	0.05	0.03	0.07	0.65	0.08
q7	0.19	0.04	0.23	0.05	0.41	0.08	-0.13	0.07

q1: I haven't been able to behave socially.

q2: I haven't been able to maintain the quality of my work.

q3: I have had trouble thinking clearly.

q4: I have taken more rests during my work.

q5: I have felt that my work isn't going well.

q6: I haven't been able to make rational decisions.

q7: I haven't been proactive about my work.

^aItems:

TABLE 3 Effect size of differential item functioning (DIF) by different language versions

Item ^a	Japanese (n = 1000)		Chinese (n = 400)		American (n = 300)	
	effect size	SE	effect size	SE	effect size	SE
q1	-0.30	0.05	0.42	0.07	0.25	0.08
q2	0.04	0.05	-0.16	0.06	0.13	0.08
q3	-0.08	0.05	0.19	0.06	-0.07	0.08
q4	0.23	0.05	-0.05	0.07	-0.52	0.08
q5	0.04	0.05	-0.16	0.06	0.16	0.08
q6	0.05	0.05	-0.34	0.06	0.41	0.08
q7	0.06	0.05	0.12	0.07	-0.34	0.08

Effect size refers to logit differences in Rasch difficulty estimates between target subjects and total subjects.

Effect size more than 0.50 logit in absolute value indicates presence of DIF.

q1: I haven't been able to behave socially.

q2: I haven't been able to maintain the quality of my work.

q3: I have had trouble thinking clearly.

q4: I have taken more rests during my work.

q5: I have felt that my work isn't going well.

q6: I haven't been able to make rational decisions.

q7: I haven't been proactive about my work.

^aItems:

adaptation is a process that ensures equivalence in meaning, with equivalence comprising several components, including conceptual equivalence and item equivalence.^{33,34} This process was reflected in the following steps used in the present study: initial translation, synthesis/reconciliation of the translations, back translation, expert committee review, and pretesting.^{23,35} Following cross-cultural adaptation, cross-cultural validation is examined, in which particular scrutiny is placed on measurement invariance. This refers to target populations with comparable disease severity; that is, scores obtained using the original and cross-culturally adapted versions are the same.¹⁹ Such cases would not exhibit DIF. The Rasch model is a well-known method for detecting DIF.

DIF was negligible for all seven items in the Chinese version and six items in the English version of WFun. This suggests that there are few linguistic or cultural biases affecting the question items for Chinese and American subjects. We only identified DIF in the item “I have taken more rests during my work” in the English version. This indicates that American respondents had a higher tendency to affirm this item than Japanese and Chinese respondents. Although the reason for this is unclear, American respondents may be more prone to interruptions under poor health conditions because of higher job control than subjects from Japan and China.

Similarly, the lack of DTF, which examines discrepancies between whole tests, was consistent, indicating that there was

Total score	Japanese (n = 1000)		Chinese (n = 400)		American (n = 300)	
	measure	SE	measure	SE	measure	SE
7	-5.00	1.80	-5.83	1.85	-4.40	1.82
8	-3.73	1.00	-4.56	1.05	-3.20	1.00
9	-2.96	0.70	-3.77	0.78	-2.52	0.71
10	-2.47	0.60	-3.26	0.66	-2.11	0.58
11	-2.10	0.50	-2.86	0.60	-1.81	0.51
12	-1.79	0.50	-2.52	0.56	-1.57	0.47
13	-1.53	0.50	-2.22	0.54	-1.37	0.44
14	-1.29	0.40	-1.94	0.52	-1.19	0.41
15	-1.08	0.40	-1.67	0.51	-1.02	0.40
16	-0.87	0.40	-1.41	0.51	-0.87	0.39
17	-0.69	0.40	-1.15	0.50	-0.72	0.38
18	-0.50	0.40	-0.90	0.50	-0.58	0.38
19	-0.33	0.40	-0.65	0.50	-0.43	0.38
20	-0.16	0.40	-0.39	0.51	-0.29	0.38
21	0.01	0.40	-0.13	0.51	-0.15	0.38
22	0.18	0.40	0.13	0.52	0.00	0.39
23	0.35	0.40	0.40	0.53	0.16	0.40
24	0.52	0.40	0.68	0.53	0.32	0.41
25	0.70	0.40	0.97	0.54	0.49	0.42
26	0.89	0.40	1.27	0.55	0.68	0.44
27	1.09	0.40	1.59	0.57	0.89	0.46
28	1.30	0.40	1.91	0.58	1.11	0.49
29	1.53	0.40	2.26	0.59	1.36	0.51
30	1.79	0.50	2.62	0.62	1.64	0.55
31	2.09	0.50	3.02	0.65	1.97	0.60
32	2.45	0.60	3.47	0.70	2.36	0.66
33	2.93	0.70	4.03	0.81	2.88	0.78
34	3.70	1.00	4.87	1.07	3.68	1.05
35	4.96	1.80	6.17	1.86	4.96	1.85

TABLE 4 Rasch measurements according to raw scores by language

no DTF among the three versions. The negligible DTF reflects the lack of DIF in the Chinese version, and the presence of DIF in only one item in the American version. These findings indicate that the Japanese, Chinese, and English versions of WFun are comparable.

The number of foreign workers in the Japanese labor market is increasing as a result of the declining birth rate and aging population in Japan. Moreover, the number of foreign workers employed overseas is increasing due to globalization of economic activities. Against this backdrop, health management of not only Japanese but also foreign workers is becoming increasingly important in many Japanese companies. However, given that health care, medical delivery systems, awareness of health, and the scope of safety considerations differ between Japan and other countries, health management based on diagnosis

according to a disease name and medical examination results is ineffective. We propose that management based on the degree of difficulty in conducting work, rather than a disease name or examination results, may be useful as a screening tool for health management in the global labor market.

Several limitations of this study warrant mention. First, the subjects of this study were not representative of the examined countries because sampling was conducted using an internet survey. Given that verification of DIF depends on the sample, the results of this study may not necessarily reflect those of a representative group. However, sampling in this study was non-systematic or haphazard, and, therefore, does not represent a specific group.

Second, the present study was not limited to foreigners working in Japanese companies. Foreigners

who work in Japanese companies may experience different socioeconomic conditions to those who work in non-Japanese companies. If there is a cross-cultural difference, there may be differences in interpretability or understanding of their respective languages between foreigners who work in Japanese companies and those who work in non-Japanese companies. However, there is no reason to assume this. Furthermore, the Rasch model assumes no sample dependence for test items of measures produced by the model, a property called “specific objectivity”.^{26,36,37} In fact, we examined this property in the development process and found that estimates of item difficulty were consistent between subgroups of different sex, age, income, and job type, as well as different companies.¹⁵ Nonetheless, to further verify comparability with Japanese workers, future studies should conduct a survey targeting foreigners who work for Japanese companies.

Third, validation of the English version among an American population does not guarantee its validity among other English-speaking populations.

In conclusion, there was no DIF in the Chinese or English version of WFun except for one item in the English version. Likewise, there was no DTF in either the Chinese or English version. This study suggests that results from the English, Chinese, and Japanese versions of WFun have good comparability.

ACKNOWLEDGMENTS

This study was partly supported by UOEH Research Grant for Promotion of Occupational Health

DISCLOSURE

Approval of the research protocol: N/A. Informed consent: N/A. Registry and the registration no. of the study/trial: N/A. Animal studies: N/A. Conflicts of interest: The authors have no conflict of interest.

AUTHOR CONTRIBUTIONS

YF collected and analyzed the data and led the writing. NL, OC, MO, TI, and TK supported writing.

ORCID

Yoshihisa Fujino  <https://orcid.org/0000-0002-9126-206X>

Ning Liu  <https://orcid.org/0000-0002-5257-2719>

Odgerel Chimed-Ochir  <https://orcid.org/0000-0003-0159-3701>

Tomohiro Ishimaru  <https://orcid.org/0000-0002-9334-1423>

Tatsuhiko Kubo  <https://orcid.org/0000-0001-8822-9181>

REFERENCES

- Brooks A, Hagen SE, Sathyanarayanan S, Schultz AB, Edington DW. Presenteeism: critical issues. *J Occup Environ Med.* 2010;52:1055-1067.
- Johns G. Presenteeism in the workplace: a review and research agenda. *J Organ Behav.* 2010;31:519-542.
- Burton WN, Conti DJ, Chen CY, Schultz AB, Edington DW. The economic burden of lost productivity due to migraine headache: a specific worksite analysis. *J Occup Environ Med.* 2002;44:523-529.
- Edington DW. Health and productivity. In: McCunney RJ, ed. *A practical approach to occupational and environmental medicine.* London, UK: Lippincott, Williams & Wilkins; 2003:140-152.
- Lerner D, Amick BC III, Lee JC, et al. Relationship of employee-reported work limitations to work productivity. *Med Care.* 2003;41:649-659.
- Loeppke R, Hymel PA, Lofland JH, et al; American College of Occupational and Environmental Medicine. Health-related workplace productivity measurement: general and migraine-specific recommendations from the ACOEM Expert Panel. *J Occup Environ Med.* 2003;45:349-359.
- Dean BB, Aguilar D, Barghout V, et al. Impairment in work productivity and health-related quality of life in patients with IBS. *Am J Manag Care.* 2005;11:S17-26.
- Adler DA, McLaughlin TJ, Rogers WH, Chang H, Lapitsky L, Lerner D. Job performance deficits due to depression. *Am J Psychiatry.* 2006;163:1569-1576.
- Loeppke R, Taitel M, Richling D, et al. Health and productivity as a business strategy. *J Occup Environ Med.* 2007;49:712-721.
- Schultz AB, Chen CY, Edington DW. The cost and impact of health conditions on presenteeism to employers: a review of the literature. *Pharmacoeconomics.* 2009;27:365-378.
- Lerner D, Amick BC III, Rogers WH, Malspeis S, Bungay K, Cynn D. The work limitations questionnaire. *Med Care.* 2001;39:72-85.
- Koopman C, Pelletier KR, Murray JF, et al. Stanford presenteeism scale: health status and employee productivity. *J Occup Environ Med.* 2002;44:14-20.
- Wahlqvist P, Carlsson J, Stalhammar NO, Wiklund I. Validity of a work productivity and activity impairment questionnaire for patients with symptoms of gastro-esophageal reflux disease (WPAI-GERD)—results from a cross-sectional study. *Value Health.* 2002;5:106-113.
- Kessler RC, Barber C, Beck A, et al. The world health organization health and work performance questionnaire (HPQ). *J Occup Environ Med.* 2003;45:156-174.
- Fujino Y, Uehara M, Izumi H, et al. Development and validity of a work functioning impairment scale based on the Rasch model among Japanese workers. *J Occup Health.* 2015;57:521-531.
- Nagata T, Fujino Y, Saito K, et al. Diagnostic accuracy of the work functioning impairment scale (WFun): a method to detect workers who have health problems affecting their work and to evaluate fitness for work. *J Occup Environ Med.* 2017;59:557-562.
- Makishima M, Fujino Y, Kubo T, et al. Validity and responsiveness of the work functioning impairment scale (WFun) in workers with pain due to musculoskeletal disorders. *J Occup Health.* 2018;60:156-162.
- Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and

- definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63:737-745.
19. Henrica CWd. *Measurement in medicine: a practical guide*. Cambridge, UK: Cambridge University Press; 2011.
 20. Brodersen J, Meads D, Kreiner S, Thorsen H, Doward L, McKenna S. Methodological aspects of differential item functioning in the Rasch model. *J Med Econ*. 2007;10:309-324.
 21. Wilson M. *Constructing measures: an item response modeling approach*. New York, NY: Psychology Press; 2005.
 22. Mullen MR. Diagnosing measurement equivalence in cross-national research. *J Int Bus Stud*. 1995;26:573-596.
 23. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 1976;2000(25):3186-3191.
 24. Rasch G. An item analysis which takes individual differences into account. *Br J Math Stat Psychol*. 1966;19:49-57.
 25. Andrich D. *Rasch models for measurement*. Newbury Park, UK: Sage; 1988.
 26. Bond TG, Fox CM. *Applying the Rasch model: fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates; 2007.
 27. Linacre JM. Sample size and item calibration stability. *Rasch Measurement Transactions*. 1994;7:328.
 28. Tristan A. An adjustment for sample size in DIF analysis. *Rasch Measurement Transactions*. 2006;20:1070-1071.
 29. Munkholm M, Berg B, Lofgren B, Fisher AG. Cross-regional validation of the school version of the assessment of motor and process skills. *Am J Occup Ther*. 2010;64:768-775.
 30. Wright B, Stone M. *Measurement essentials*. Wilmington, DE: Wide Range; 1999.
 31. Pae T-I, Park G-P. Examining the relationship between differential item functioning and differential test functioning. *Language Testing*. 2006;23:475-496.
 32. ShROUT PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420-428.
 33. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J Clin Epidemiol*. 2015;68:435-441.
 34. Flaherty JA, Gaviria FM, Pathak D, et al. Developing instruments for cross-cultural psychiatric research. *J Nerv Ment Dis*. 1988;176:257-263.
 35. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46:1417-1432.
 36. Wright BD, Masters GN. *Rating scale analysis*. Chicago, IL: Mesa Press; 1982.
 37. Nering ML, Ostini R. *Handbook of polytomous item response theory models*. New York, NY: Routledge; 2010.

How to cite this article: Fujino Y, Liu N, Chimed-Ochir O, Okawara M, Ishimaru T, Kubo T. Cross-cultural validation of the work functioning impairment scale (WFun) among Japanese, English, and Chinese versions using Rasch analysis. *J Occup Health*. 2019;61:464-470. <https://doi.org/10.1002/1348-9585.12072>