Sports Medicine and Health Science 2 (2020) 153–158

Contents lists available at ScienceDirect

Sports Medicine and Health Science

journal homepage: www.keaipublishing.com/en/journals/sports-medicine-and-health-science/

Original Research

Validity and reliability of a Chinese rating of perceived exertion scale in young Mandarin speaking adults

Weiyang Ding [a,b], Tongjian You [c], Philimon N. Gona [c], Laurie A. Milliken [c,*]

[a] Northern Jiangsu People's Hospital, Yangzhou, Jiangsu, China
[b] Yangzhou University, Clinical Medical College, Yangzhou, Jiangsu, China
[c] Department of Exercise and Health Sciences, University of Massachusetts Boston, Boston, MA, USA

ARTICLE INFO

Keywords:
Borg 6–20 RPE Scale
Perceived exertion
Psychophysiological tool
Translation

ABSTRACT

Background: The validity and reliability of the Borg 6–20 rating of perceived exertion (RPE) scale has not been tested among Chinese people from Mainland China. The purpose of this study was to test: 1) The validity of Leung Chinese version and Wang Chinese version of the Borg 6–20 RPE scale; 2) The reliability of Wang Chinese version RPE scale; and 3) The agreement of these two Chinese versions of the RPE scale among young healthy adults from Mainland China.

Methods: A total of 26 subjects (11 males, 15 females; age 22.7 ± 3.0 yrs) volunteered to participate. They performed one ($n = 3$), two ($n = 14$), or three trials ($n = 9$) of the Bruce treadmill protocol test within 9.0 ± 5.1 days (validation trials), and 30.4 ± 27.9 days (reliability trials). Power output, heart rate, oxygen consumption, and RPE were recorded.

Results: RPE was significantly correlated with power output (Leung version $rs \geq 0.75$, Wang version $rs \geq 0.73$), heart rate (HR) (Leung version $rs \geq 0.84$, Wang version $rs \geq 0.87$), and oxygen consumption (VO_2) (Leung version $rs \geq 0.80$, Wang version $rs \geq 0.81$) (all $p < 0.01$). The overall test-retest interclass correlation was 0.94 ($p < 0.01$). No significant differences in correlations (RPE against power output, HR and VO_2) between trials existed for the reliability tests of Wang version scale. No significant differences in correlations (RPE against power output, HR and VO_2) between the two Chinese versions of RPE scale existed.

Conclusion: Both Chinese RPE scales are valid among young healthy Chinese mandarin speaking adults. The Wang scale is reliable, and the Leung and Wang scales show superior agreement with each other.

Introduction

Borg's 6–20 rating of perceived exertion scale (RPE) is a valid and reliable psychophysiological tool to measure perceptions of exertion during exercise and has been translated into different languages.[1] In Mainland China, this RPE scale has been popularly used in exercise and health fields but its Chinese translation varies. The Standards for Educational and Psychological Testing[2] stated that whenever a test is translated from one language or dialect to another, its reliability and validity should be established in the target linguistic group. Because this testing has not been performed in a linguistic group from Mainland China who speak Mandarin Chinese, the RPE scale should not be widely adopted in China until its validity and reliability are confirmed for this linguistic group.

The reasons to test a Cantonese version[3] and a Mandarin version[4] of the RPE scale are as follows. Leung translated the Borg RPE scale into Chinese and established the validity and reliability of the scale among Cantonese speaking children,[3] young adults,[5] and older adults[6] in Hong Kong. Chen,[7] using the Leung version of the RPE scale, assessed its validity among young Taiwanese men. But Leung's version of this scale has not been tested on Mandarin speaking people from Mainland China. Furthermore, these dialects spoken in Hong Kong and Taiwan differ from those spoken in Mainland China making a validation necessary.

Wang[4] translated the Borg RPE scale from English into another Chinese version as a part of the translation of ACSM's Guidelines for Exercise Testing and Prescription (8th edition), which was published in Mainland China. There is no evidence shown that this Chinese version RPE scale has been tested for validity and reliability in a Mandarin speaking Chinese population. Besides other unofficial varied Chinese translations of the Borg's RPE scale, there is evidence[8] showing even Wang's team is not

* Corresponding author. Department of Exercise and Health Sciences, University of Massachusetts Boston, 100 Morrissey Blvd, Boston, MA, 02125, USA.
E-mail address: laurie.milliken@umb.edu (L.A. Milliken).

https://doi.org/10.1016/j.smhs.2020.08.001
Received 21 March 2020; Received in revised form 13 May 2020; Accepted 19 August 2020
Available online 28 August 2020

conducting their own translated version of this scale in a consistent format. For all the reasons above, there is an urgent need to perform this study, not only for the matter of fulfilling the methodologic procedure, but also for Mandarin-speaking Chinese people who will gain the benefits from using this standard tool.

Because the Leung version of the RPE scale has not been tested among people from Mainland China and the Wang version of the scale has not been tested at all, the purpose of this study was: 1) to assess the validity of Leung's[3] and Wang's[4] Chinese versions of the Borg's 6–20 RPE scale among healthy Mandarin speaking Chinese adults from Mainland China; 2) to assess the reliability of Wang's[4] Chinese version Borg's 6–20 RPE scale; and 3) to assess the level of agreement between these two scales.

Methods

Participants

Flyers and emails were used to solicit potential participants and were distributed to Chinese communities in the Boston area. Participants were included if they were healthy male and female volunteers who were born and lived in mainland China for more than 15 years, whose native language was Chinese, who were between 18 and 30 years old, and who were in the low cardiovascular disease (CVD) risk category (<2 risk factors).[9] Participants were excluded if they were not fluent in Mandarin Chinese (Pu Tong Hua), had physical disabilities, and were taking medications that effected HR and blood pressure. This study was approved by the University of Massachusetts Boston Institutional Review Board. All study procedures were explained to each participant and written informed consent was obtained from all participants before baseline measurements started. A copy of the signed consent form was given to each participant.

Procedures

Two maximal exercise testing sessions took place within an average of 9 days between two validation tests (range 6–21 days, SD 5.1 days), and an average of 30.4 days between two reliability tests (range 7–74 days, SD 27.9 days). Assessments were carried out at the Exercise Physiology lab, Department of Exercise and Health Sciences, University of Massachusetts Boston. Participants were asked to refrain from strenuous physical activity for 24 h prior to testing, and to abstain from eating 3 h before the visit. During the initial visit, participants signed the informed consent documents, were familiarized with equipment, and underwent randomization for the order of performing the tests using the two RPE scales. As a result of this randomization, thirteen used the Wang[4] version RPE scale during the initial visit and eleven used the Leung[3] version RPE scale on the initial visit. The opposite scale was used on the follow-up visit.

Throughout the testing procedures all verbal instructions were given in Mandarin Chinese. Before the assessment, subjects read instructions for the standardized Borg RPE scale that had been translated into Mandarin Chinese. During every test the RPE scale was positioned in front and in reach of each participant and was carefully reviewed. Participants were instructed to point to the number on the scale that indicated their perceived effort. It was emphasized to the participants that the rating was dependent upon their overall, whole-body perception of exertion, and was not limited to fatigue experienced in any one area in the legs, chest, or cardiovascular system. Following the verbal explanation, participants were given approximately 10 min to study the scale and to ask any questions.

The Bruce Treadmill Protocol[10] was used to conduct all maximal exercise tests. In each session, participants began at 0% grade and a 1.7 mph walking pace for 3 min as a warm-up. During the first stage (minutes 1 to 3) of the test, the participant walked at a 1.7 mph pace at 10% grade. At the start of the second stage (minutes 4 to 6), the grade increased by 2% and the speed increased to 2.5 mph (67 m/min). In each subsequent stage of the test, the grade was increased by 2% and the speed by either 0.8 or 0.9 mph (21.4 or 24.1 m/min). Participants continued exercising to the point of volitional fatigue. This protocol was programed and input into the ParvoMedics computer protocol system (ParvoMedics Trueone 2400, Sandy, UT, USA), which automated the treadmill.

Measurement

Height (m) and body weight (kg) were measured at the beginning of each session, using standard laboratory procedures (Seca 703 Waist-High Digital Scale, Chino, CA). During each maximal treadmill test, the participant's HR was measured by telemetry (Polar Electro, Kempele, Finland), and maximal oxygen consumption (VO_{2max}) was directly measured by gas analysis of expired air using a metabolic cart (ParvoMedics Trueone 2400, Sandy, UT, USA). The True One 2400 system was calibrated against medical grade gasses (ParvoMedics, Sandy, UT, USA), and average VO_2 values were computed over 5 s intervals. The system also calculated respiratory exchange ratios (RER; volume expired CO_2 divided by volume expired O_2), which were used as an indicator for achieving VO_{2max}. An RER is or above 1.15 was used to indicate that VO_{2max} had been reached. The VO_2, RPE and HR responses were obtained in the last 10 s of each minute. If the participant did not reach 10 s of the 3-min stage, the last 10 s of VO_2, RPE and HR values were determined as the maximal values. All participants exercised until volitional fatigue. Graded exercise testing protocol and termination criteria followed the ACSM guidelines.[9]

Sample size estimates

In order to determine how many participants were needed to assess the validity of these RPE scales, we examined other studies where RPE was compared to physiological responses.[5,7] Assuming an $r = 0.60$ between RPE and objective measures, we needed 18 participants to achieve a power of 80%, $p < 0.05$ (for a two-sided test). We conservatively estimated, with the participant dropout rate at 10%, that the sample size of 20 participants was adequate for testing the validities of RPE-Leung and RPE-Wang.[11]

In order to determine how many participants were needed to assess the reliability of these RPE scales, we examined Leung's study.[5] Assuming an $R = 0.80$, we needed 9 participants to achieve a power of 80%, $p < 0.05$ (for a two-sided test). We conservatively estimated, with a dropout rate at 10%, that the sample size of 10 participants was adequate for testing the validities of RPE-Leung and RPE-Wang.[11]

Statistical analyses

Pearson correlation coefficients were computed to quantify the concurrent validity of both Chinese versions of the RPE scale[3,4] against the measures of objective exercise intensity (power output, HR, and VO_2) using two statistical procedures (Methods 1 and 2).[3] For Method 1, Pearson *r* was computed for each participant for RPE versus each objective measure of intensity (power output, HR and VO_2), and then the mean of all Pearson *r*'s was calculated for each objective measure. For Method 2, Pearson *r* was computed from the simultaneous analysis of all subjects' data to find a single overall linear correlation coefficient for each objective measure.

The test-retest reliability of the Wang version RPE scale across two trials was quantified using three procedures. First, the intra-class correlations (*ICC*)[12] was calculated. Second, the 95% limits of agreement procedure from Bland - Altman[13,14] was used to examine the RPE values recorded for each exercise intensity. Third, Pearson correlation coefficients were calculated using Methods 1 and 2 described above. The Fisher z-transformation test was used to determine if the *r*'s differed significantly between reliability trials.

The agreement between the Wang version and the Leung version of the RPE scale was assessed by four procedures. First, *ICCs*[12] of the RPE

Table 1
Demographic characteristics for participants.

	All ($n = 26$)	Men ($n = 11$)	Women ($n = 15$)
Age, yr	22.7 (3.0)	22.0 (2.6)	23.2 (3.0)
Mass, kg	64.9 (13.9)	70.1 (17.4)	61.2 (8.3)
Height, m	1.7 (0.1)	1.7 (0.1)	1.6 (0.1)
BMI, kg/m^2	22.9 (3.9)	23.2 (4.7)	22.8 (2.9)
Been in the US., years	2.1 (1.8)	2.5 (2.1)	1.8 (1.4)
Days between Leung and Wang Trials	9.0 (5.1)	9.3 (5.8)	8.8 (4.2)
Days between Wang Trials 1 and 2	30.4 (27.9)	20.5 (16.4)	50.3 (30.7)

Values are mean (standard deviation).

values from the two trials to validate the Leung and Wang versions of the RPE scale were obtained. Second, the 95% limits of agreement procedure from Bland-Altman[13,14] was used to examine the RPE values recorded for each exercise intensity. Third, the Fisher *z*-transformation test was used to determine if the *r*'s differed significantly between the two trials for validating the Leung version and the Wang version of the RPE scale. Fourth, paired *t*-tests were utilized to compare the paired differences of heart rate and VO_2 at the RPE scale at 13, 15, 17, and 19, which are the major differences between the two RPE scales.

All statistical procedures were conducted using the SPSS 24.0 for Windows, and the type 1 error rate was set at $p \leq 0.05$.

Results

Demographics

Table 1 shows the demographic information for the participants. This study included 26 participants (11 men and 15 women) who had been in the United States for 2.1 ± 1.8 years. The participants were relatively young adults. The mean age for men was 22 ± 2.6 years and the mean age for women was 23.2 ± 3.0 years. At an average BMI of 22.9 ± 3.9, the sample was of normal weight. Participants were from seven providences presented: Shanxi (1), Henan (3), Jiangsu (3), Gansu (2), Fujian (4), Guangdong (3), Yunnan (1); one autonomous region, Tibet (1), and four direct-controlled municipalities, Beijing (3), Tianjin (2), Shanghai (2), Chongqing (1).

One male and one female could not attend their second visits to this study because of schedule conflicts. One 21-year-old female, in her second trial, had her highest heart rate at 174 bpm ($HR_{max} = 199$ bpm), which was much lower than her first trial. She confirmed that she did not endeavor to try her best during this test, so her data were excluded from the analysis. For these three tests, single test data were used to compare RPE to objective measures of intensity but could not be used for between scale comparisons.

Twenty-four subjects (male $n = 11$, female $n = 13$) participated in validity study, which was more than we estimated for an 80% power. Eleven subjects (male $n = 6$, female $n = 5$) participated in the reliability study on the Wang version of the RPE scale (2010), which also should have provided adequate power. Fifty-eight tests were performed in total. All tests reached RER = 1.15. Two tests terminated at RPE 17. Fifty-six tests reached RPE 18 or above. All tests reached Stage 3.55% tests reached Stage 4.12% tests reached Stage 5. Four tests reached 85–89% of HR_{max} ($HR_{max} = 220$ - Age). Nineteen tests reached 90–94% of HR_{max}. Twenty-seven tests reached 95–99% of HR_{max}. Eight tests reached 100% or higher of HR_{max}.

Validity

Both Chinese versions of the RPE scale[3,4] were found to be valid with strong and significant Pearson correlation coefficients between the measures of objective exercise intensity (power output, HR, and oxygen consumption) using two statistical procedures (Method 1 and 2)[5] (Table 2). For Method 1, significant ($p < 0.001$) Pearson correlation coefficients were found where RPE values were correlated with power output (Leung version ≥0.98; Wang version≥0.97), HR (Leung version≥0.97; Wang version≥0.97), and oxygen consumption (Leung≥0.98; Wang≥0.99). For Method 2, significant ($p < 0.001$) Pearson correlation coefficients were found where RPE values were significantly linearly correlated with power output (Leung version≥0.75; Wang version≥0.73), HR (Leung version≥0.84; Wang version≥0.87), and oxygen consumption (Leung≥0.80; Wang≥0.81).

Table 2
Pearson correlations between ratings of perceived exertion and objective measures of exercise intensity for the wang and leung scales.

	Method 1		Method 2	
	Wang Scale (2010)	Leung Scale (2002)	Wang Scale (2010)	Leung Scale (2002)
All				
RPE vs. Power	0.97	0.98	0.73	0.75
RPE vs. HR	0.97	0.97	0.87	0.84
RPE vs. VO_2	0.98	0.99	0.81	0.80
Men ($n = 10$)				
RPE vs. Power	0.97	0.98	0.78	0.79
RPE vs. HR	0.98	0.98	0.90	0.87
RPE vs. VO_2	0.99	0.99	0.87	0.84
Women ($n = 11$)				
RPE vs. Power	0.98	0.99	0.78	0.83
RPE vs. HR	0.97	0.98	0.81	0.82
RPE vs. VO_2	0.97	0.99	0.81	0.89

* All correlations significant at $p < 0.001$; HR is heart rate; RPE is rating of perceived exertion; VO_2 is maximal oxygen consumption; Method 1is the mean of individual correlations; Method 2 is the simultaneous analysis of all data.

Reliability

The Wang translation of the RPE scale was assessed for reliability in three different ways (ICCs, Bland-Altman analyses, and Fisher z-transformation test). First, the overall test-retest ICC R of the RPE values from the two trials was 0.94 ($p < 0.01$), which means the correlation of Trial 1 and Trial 2 is highly correlated.

The 95% limits of agreement procedure from Bland-Altman were used to examine the RPE values recorded for each exercise intensity. This technique requires the calculation of the mean difference (bias) of RPE values between Trial 1 and Trial 2 and ± 1.96 x SD of these differences (the 95% limits). Bland-Altman plots were created for each stage where at most 1 data point was outside of the 95% limits of agreement. This indicates that Trial 1 and Trial 2 of the Wang scale agree well. Fig. 1a. shows an example of a Bland-Altman plot for stage 2.

Pearson correlations were calculated between RPE and objective measures (power output, HR and VO_2) by Method 1 and Method 2 for each reliability trial. For Method 1, significant ($p < 0.001$) Pearson correlation coefficients were found where RPE values were highly positively linearly correlated with power output (Trial 1 ≥ 0.97; Trial 2 ≥ 0.98), HR (Trial 1 ≥ 0.98; Trial 2 ≥ 0.98), and oxygen consumption (Trial 1 ≥ 0.99; Trial 2 ≥ 0.99) (Table 3). For Method 2, significant ($p < 0.001$) Pearson correlation coefficients were found where RPE values were positively linearly correlated with power output (Trial 1 ≥ 0.77; Trial 2 ≥ 0.79), heart rate (Trial 1 ≥ 0.88; Trial 2 ≥ 0.87), and oxygen consumption (Trial 1 ≥ 0.82; Trial 2 ≥ 0.83) (Table 3). Using the Fisher z-transformation test, no significant differences were found between pairs of r's for trial 1 and trial 2 (Method 1 p-values ranged from 0.73 to 0.76; Method 2 p-values were 0.93 for all).

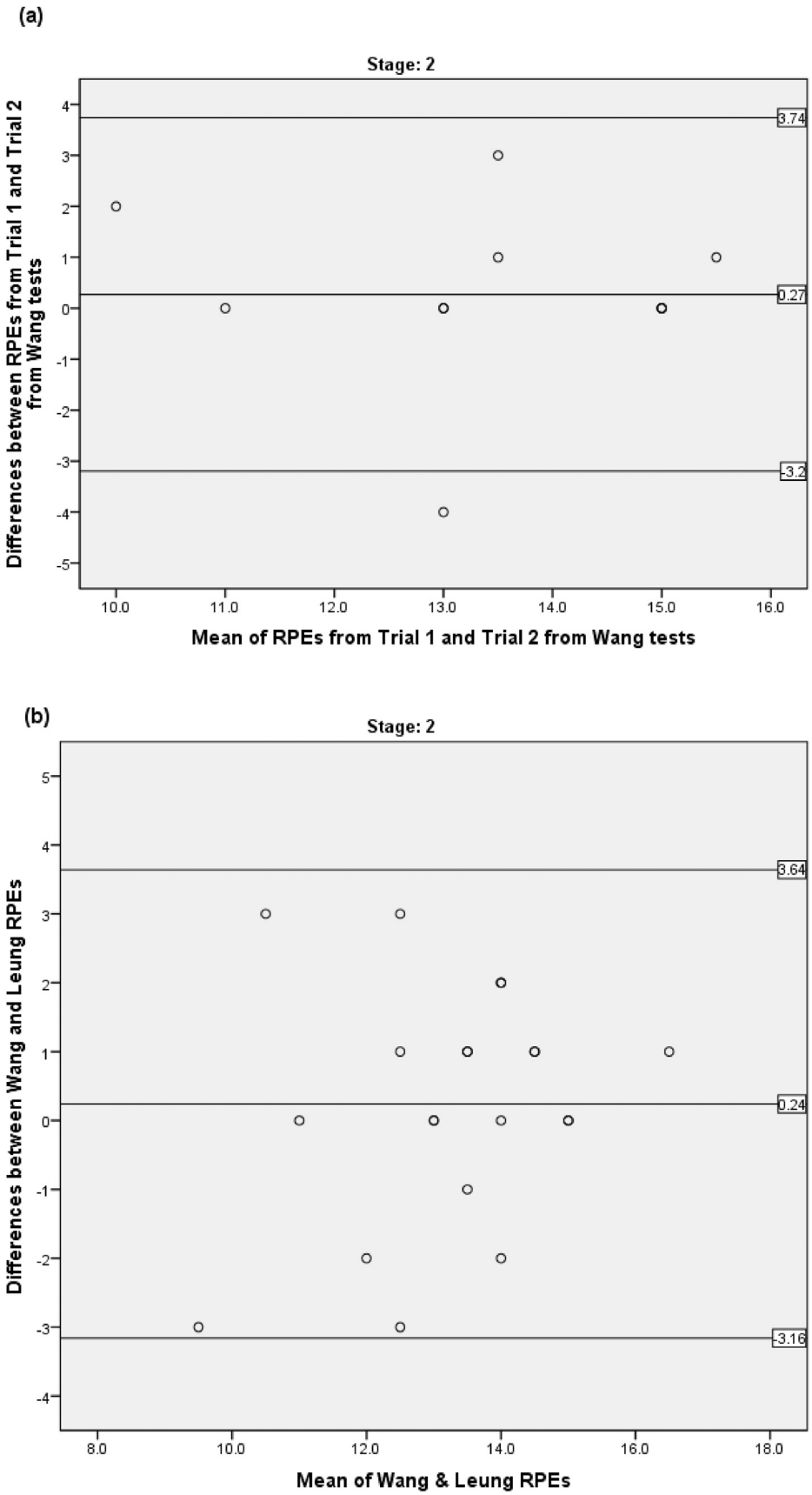

Fig. 1. (a) Bland Altman Plots showing the agreement of trial 1 and trial 2 of the Wang version of the RPE scale and (b) the agreement of the Wang and the Leung versions of the RPE scales for Stage 2 of the treadmill tests.

Table 3
Pearson correlation Between Ratings of Perceived Exertion And Objective Measures Of Exercise Intensity For Reliability Trials Of The Wang Scale.

	Method 1		Method 2	
	Trial 1	Trial 2	Trial 1	Trial 2
All ($n = 11$)				
RPE vs. Power	0.97	0.98	0.77	0.79
RPE vs. HR	0.98	0.98	0.88	0.87
RPE vs. VO_2	0.99	0.99	0.82	0.83
Men ($n = 6$)				
RPE vs. Power	0.96	0.97	0.76	0.85
RPE vs. HR	0.98	0.98	0.91	0.93
RPE vs. VO_2	0.99	0.99	0.85	0.90
Women ($n = 5$)				
RPE vs. Power	0.97	0.98	0.92	0.88
RPE vs. HR	0.97	0.98	0.86	0.77
RPE vs. VO_2	1.00	0.99	0.86	0.87

* All correlations significant at $p < 0.001$; HR is heart rate; RPE is rating of perceived exertion; VO_2 is maximal oxygen consumption; Method 1is the mean of individual correlations; Method 2 is the simultaneous analysis of all data.

Agreement

The agreement of the Leung version and the Wang version of the RPE scales were found by using four procedures (ICCs, Bland-Altman analyses, Fisher z transformation tests, and Paired t-tests). First, the ICC of the RPE values from the two trials to validate Leung version and Wang version RPE scales were obtained. The overall test-retest ICC R for the Leung and Wang tests was 0.96 ($p < 0.05$).

The 95% limits of agreement procedure from Bland - Altman were used to examine the RPE values recorded for each exercise intensity. Bland-Altman plots were created for stages 1–3 where at most 1 data point was outside of the limits of agreement. Fig. 1b. shows an example of a Bland-Altman plot for Stage 2.

The Fisher z-transformation test was utilized to test for differences between the Pearson correlation coefficients calculated from Methods 1 and 2 from Leung version versus the Wang version of the RPE scale. No significant differences were found between the r's from the Wang tests versus the Leung tests (Method 1 p-values ranged from 0.61 to 0.64; Method 2 p-values ranged from 0.75 to 0.84).

Paired t-tests were utilized to compare the paired differences in HR and VO_2 at RPE's of 13, 15, 17, and 19 between the two trials. Only those individuals who selected these RPE values during their tests were included in these analyses. No differences existed between scales for HR or VO_2 except for heart rate at RPE 17 ($n = 3$, $t = -6.928$; $p = 0.02$) (Table 4).

Discussion

In this study, RPE values were significantly correlated with power output, heart rate, and oxygen consumption for both Leung version and Wang version scales. The findings were consistent with the findings of the Leung study[5] for validity of their Cantonese-translated RPE scale among Hong Kong adults. Leung[5] found their Cantonese version RPE scale was shown to be a valid psychophysiological tool to measure perceptions of exertion during controlled cycle ergometer exercise by Hong Kong young adults. Their significant Pearson correlation coefficients between RPE and objective measures (power output, HR and VO_2) (Method 1: the range of *r*'s 0.91–0.97; Method 2: the range of *r*'s 0.69–0.75) were similar in magnitude to our findings (Method 1: the range of *r*'s 0.97–0.99; Method 2: the range of *r*'s 0.73–0.87).

Although a treadmill protocol was used versus a cycle protocol that Leung used, this study found similar maximal testing results. Treadmills have been used to test the validity and reliability of RPE scale in other studies.[14,15] This protocol is acceptable and the test was well tolerated and appropriate for the sample. These results can be generalized to this population, and indicate that both the Leung and Wang versions of the Chinese language RPE scale are valid and reliable psychophysiological tools to measure perceptions of exertion during controlled treadmill exercise by young healthy Mandarin speaking Chinese.

In this present study, the Wang version RPE scale was found to be reliable similar to the Leung version RPE scale reliability assessments which were studied among Cantonese speaking children,[3] adults,[5] and older adults[6] in Hong Kong. In addition, Lamb[14] stated that the "95% limits of agreement" technique[16] is a better means of assessing reliability of RPE scale than traditional indicators such as Pearson correlation coefficient and ICC. These analyses were performed and found good agreement between trials indicating high reliability. No more than 1 data point fell outside of the 95% limits of agreement for the stages plotted. There was little bias and no influence of exertion level on the difference between trials. This indicates that the Wang version RPE scale is a reliable psychophysiological tool to measure perceptions of exertion during controlled treadmill exercise by young healthy Mandarin speaking Chinese.

The innovation of this study was to examine the agreement between the Leung version and the Wang version of the RPE scale. Because neither the Leung nor the Wang scale was validated in Mandarin speaking Chinese adults, this allowed the opportunity to validate both scales, and to assess the agreement. The major differences of Chinese verbal descriptions between these two scales are at the numbers of 13 (Somewhat Hard), 15 (Hard), 17 (Very Hard), and 19 (Extremely Hard) in the translation of the word "hard". Though Wang's translation[4] was more related to strenuous physical effort while Leung's translation[3] was more related to the meaning of "hard" in Chinese linguistic circumstances, it did not seem to be interpreted by participants differently. This may be because the meaning of the word "hard" ("困难" translation in Leung's version) is not different for those living in Hong Kong, Taiwan and Mainland China despite the different dialects spoken between these regions. This can be a reason why Leung's version has highly validity for Mandarin-speaking adults in China as well as those speaking Cantonese in Hong Kong and Taiwanese in Taiwan. The results of paired t-tests of comparing the differences of heart rate and VO_2 at RPE scale 13, 15, 17, and 19 did not show significant differences except when comparing heart rate at scale 17. This might be caused by having a small sample size because only three people selected an RPE of 17 during their tests. Overall, there is a high agreement between these two scales. These wording differences do not seem to matter presumably because people

Table 4
Average VO_2 and heart rate differences for wang versus leung versions of the rating of perceived exertion (RPE) scales at RPEs of 13, 15, 17, and 19.

RPE	Outcome	*n*	Wang Version of RPE Scale	Leung Version of RPE Scale	Mean Difference (SD)	*t* value	p-value
13	VO_2(ml/kg/min)	12	23.26	24.96	1.71 (4.48)	1.32	0.21
	HR(beat/min)	12	140.92	145.33	4.42 (18.02)	0.85	0.41
15	VO_2(ml/kg/min)	15	29.83	32.06	2.23 (4.42)	1.96	0.07
	HR(beat/min)	15	165.67	169.07	3.40 (11.39)	1.16	0.27
17	VO_2(ml/kg/min)	3	35.15	35.42	0.27 (1.26)	0.37	0.75
	HR(beat/min)	3	177.67	173.67	−4.00 (1.00)	−6.93	0.02
19	VO_2(ml/kg/min)	7	38.35	38.87	0.52 (2.55)	0.54	0.61
	HR(beat/min)	7	184.29	182.72	−1.57 (7.87)	−0.53	0.62

HR is heart rate; VO_2 is oxygen consumption; p-value is from paired t-tests.

successfully interpreted the scale in the context of the activity they were completing. Therefore, these scales can be used interchangeably in this population of Chinese adults.

From a methodological perspective, there are some limitations of this study. First, the sample size was a small convenience sample of average fitness, with an age range from 19 to 29 years and, from a university, which limits the generalizability of the findings to the general population from mainland China. No attempt was made to assess these RPE scales on those Chinese adults with medical conditions, those who were younger or older or those who were highly fit. The physical fitness level of participants in this study was varied. Researchers found that people with higher fitness levels had stronger correlations between RPE and HR, core temperature,[18] and exercise intensity.[19] This issue needs examination before these RPE scales are used in younger, older, highly fit, or diseased Chinese populations.

Second, nine out of eleven participants who were included in the reliability study for the Wang version RPE scale performed three tests in total, which might overestimate reliability due to familiarization with the scale. Third, the broader application of RPE is for regulation of daily exercise training intensity and the current study does little to address that. Fourth, there are controversies to test reliability with Bland-Altman test, even Lamb[14] used it to measure the reliability of the RPE scale. Other authors[1,17] have indicated that the Bland-Altman method is not adapted for perceptual scales, because reliability is not a simple and constant aspect inherent in an index but depends on the measurements obtained with it; and Bland-Altman test is probably influenced by the ratings of RPE scale, which does not use decimals.

Conclusions

This present study is the first validity assessment of the Leung[3] and Wang[4] translations, and the first reliability assessment of the Wang translation of the Borg 6–20 RPE scale among Mandarin-speaking Chinese people who are from Mainland China. The results of this study demonstrate that both Leung version and Wang version of the RPE scale are valid psychophysiological tools to measure perceptions of exertion during controlled treadmill exercise by Chinese healthy young adults. They agree with each other and can be used in place of one another. The reliability of the Wang version Chinese 6–20 RPE scale has been established. Additional research is needed to verify these findings for different age groups and for use among those with multiple health conditions.

Authors' contributions

All authors contributed significantly to the reported research and approved the final version of the submitted manuscript. The authors confirmed that the reported work is original and accurate, and that the manuscript has not been submitted elsewhere.

Submission statement

The manuscript has not been submitted elsewhere.

Ethical approval

The study was approved by the University of Massachusetts Boston Institutional Review Board and was registered under approval #2015185. Written informed consent was obtained from all participants before baseline measurements were started and a copy of the signed consent form was given to each participant.

Confilct of interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.smhs.2020.08.001.

References

1. Borg G. *Borg's Perceived Exertion and Pain Scales*. 1998. Human kinetics.
2. *Standards for Educational and Psycological Testing*. Washington, DC: American Psychological Association; 1985.
3. Leung M-L, Chung P-K, Leung R. An assessment of the validity and reliability of two perceived exertion rating scales among Hong Kong children. *Percept Mot Skills*. 2002; 95:1047–1062. https://doi.org/10.2466/pms.2002.95.3f.1047.
4. Wang Z. *ACSM's Manual for Guidelines for Exercise Testing and Prescription (8th, Chinese Version)*. Beijing, China: People's Medical Publishing House; 2010:96–97.
5. Leung RW, Leung ML, Chung PK. Validity and reliability of a Cantonese-translated rating of perceived exertion scale among Hong Kong adults. *Percept Mot Skills*. 2004; 98:725–735. https://doi.org/10.2466/pms.98.2.725-735.
6. Chung P-K, et al. A brief note on the validity and reliability of the rating of perceived exertion scale in monitoring exercise intensity among Chinese older adults in Hong Kong. *Percept Mot Skills*. 2015;121(3):805–809. https://doi.org/10.2466/29.pms.121c24x8.
7. Chen Y-L, et al. Relationships of Borg's RPE 6-20 scale and heart rate in dynamic and static exercises among a sample of young Taiwanese men. *Percept Mot Skills*. 2013; 117(3):971–982. https://doi.org/10.2466/03.08.PMS.117x32z6.
8. Luo X-j, et al. Applications of energy consumption code scale of physical activities in exercise prescription. *Journal of Beijing Sport University*. 2013;36(9):76–80.
9. *ACSM's Guidelines for Exercise Testing and Prescription*. Baltimore, MD: Willianms & Wilkins; 2009.
10. Bruce RA, et al. Exercising testing in adult normal subjects and cardiac patients. *Ann Noninvasive Electrocardiol: The Official Journal Of The International Society For Holter And Noninvasive Electrocardiology, Inc*. 1963;9(3):291–303, 1963.
11. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. second ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
12. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428. https://doi.org/10.1037//0033-2909.86.2.420.
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.
14. Lamb KL, Eston RG, Corns D. Reliability of ratings of perceived exertion during progressive treadmill exercise. *Br J Sports Med*. 1999;33:336–339. https://doi.org/10.1136/bjsm.33.5.336.
15. Morris M, et al. The validity and reliability of predicting maximal oxygen uptake from a treadmill-based sub-maximal perceptually regulated exercise test. *Eur J Appl Physiol*. 2010;109(5):983–988. https://doi.org/10.1007/s00421-010-1439-1.
16. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2):135–160. https://doi.org/10.1177/096228029900800204.
17. Coquart JBJ, Garcin M. Validity and reliability of perceptually-based scales during exhausting runs in trained male runners. *Percept Mot Skills*. 2007;104(1):254–266. https://doi.org/10.2466/pms.104.1.254-266.
18. Travlos AK, Marisi DQ. Perceived exertion during physical exercise among individuals high and low in fitness. *Percept Mot Skills*. 1996;82(2):419–424. https://doi.org/10.2466/pms.1996.82.2.419.
19. Green JM, et al. Influence of aerobic fitness on ratings of perceived exertion during graded and extended duration cycling. *J Sports Med Phys Fit*. 2007;47(1):33–39.