Research Paper

Hong Kong Physiotherapy Journal
Vol. 42, No. 2 (2022) 137–149
DOI: 10.1142/S1013702522500135

Hong Kong Physiotherapy Journal
https://www.worldscientific.com/worldscinet/hkpj

Chinese (Mandarin) translation of the incremental shuttle walk test and its validity and reliability: A cross-sectional study

Wei Qin Ang[1], Hong Ting Tan[1], Si Min Goh[1], Samantha W. Seng[1], Katherin S. Huang[1,2], Melissa Y. Chan[1] and Meredith T. Yeung[1,*]

[1]*Health and Social Sciences Cluster*
Singapore Institute of Technology, Singapore

[2]*Department of Physiotherapy*
Khoo Teck Puat Hospital, Singapore

Meredith.Yeung@SingaporeTech.edu.sg

Received 30 July 2021; Accepted 19 July 2022; Published 31 August 2022

Background/Purpose: To date, there are no published validated Chinese versions of the incremental shuttle walk test (ISWT) instructions despite its wide clinical applications. Translation of the Chinese ISWT instruction is done in an *ad-hoc* manner within the Chinese-speaking populations, affecting the test's reliability and validity since translation can differ significantly between individuals. This warrants the need for psychometric testing of such translation.
Objectives: To develop a Chinese (Mandarin) version of the ISWT instructions (ISWT-CHN) that is conceptually equivalent to the original English version (ISWT-ENG) and establish its reliability and validity.
Methods: Forward and backward translations from the ISWT-ENG were done to generate the ISWT-CHN. Face and content validity was determined during the translation process. Intra-rater and inter-rater reliability of the ISWT-CHN, construct and criterion validity were established by analysing the ISWT and the gold standard cardiopulmonary exercise test results.
Results: The Item-Content validity index (I-CVI), Scale-level-Content validity index (S-CVI), and content validity ratio (CVR) of the ISWT-CHN were 1.0. Intra-class Correlation Coefficient (ICC) for inter-rater reliability between two raters were excellent (ICC = 0.99, 95% CI 0.97–1.0, $p < 0.001$; SEM = 0.85 m, MDC = 2.35 m). The intra-rater reliability of both Raters A (ICC = 0.92, 95% CI 0.53–0.98, $p = 0.003$;

*Corresponding author.

SEM = 35 m, MDC = 97 m) and B (ICC = 0.90, 95% CI 0.76–0.96, $p < 0.001$; SEM = 32 m, MDC = 88 m) were good. In a sample of 32 healthy participants, both ISWT-CHN and ISWT-ENG instruction results showed low-positive correlations with the VO_{2max} determined from the cardiopulmonary exercise test ($r = 0.439$, $p < 0.001$; $r = 0.448$, $p < 0.001$). There is a very high correlation between ISWT-ENG and ISWT-CHN results with no statistically significant differences ($r = 0.967$, $p < 0.001$). The construct and criterion validity of the ISWT-CHN were established.

Conclusion: This study developed the ISWT-CHN and showed that it is a valid and reliable measure conceptually comparable to the ISWT-ENG. It will benefit the determination of functional exercise capacity in Chinese-speaking populations.

Key messages

- This study is aimed to develop a Chinese (Mandarin) version of the ISWT instructions.
- The ISWT Chinese translation is valid and reliable that is conceptually comparable to the original English instruction.
- The translated ISWT-Chinese instruction will enable the use of ISWT among the Chinese-speaking populations.

Keywords: Chinese translations; cross-sectional studies; incremental shuttle walk test; reproducibility of results.

Introduction

The Incremental Shuttle Walk Test (ISWT) is a commonly-used maximal field test that measures the total distance covered by the number of completed shuttles (ISWD).[1] The ISWT was first used to determine the functional exercise capacity in people with Chronic Obstructive Pulmonary Disease (COPD),[1] and was proven to be reliable and valid.[2–5] The application of ISWT was then expanded and proven to be reliable and valid in populations with other health conditions such as cardiovascular diseases,[6,7] lung cancer,[8] and peripheral arterial diseases.[9] Functional exercise capacity refers to the maximum amount of aerobic work an individual can sustain, defined by the maximal oxygen uptake ($VO_{2\,max}$). This variable is highly associated with a patient's ability to perform activities of daily living and the results are often used for exercise prescription among physiotherapists.[10] The ISWT is used to assess physical fitness and design personalised walking programmes,[11–13] as well as assess treatment outcomes and predict rehospitalisation, morbidity, and mortality rates.[14,15]

Traditionally, Cardiopulmonary Exercise Tests (CPETs) are the gold standard for measuring an individual's exercise capacity.[16] However, they require sophisticated equipment and trained staff to conduct the test and analyse its results. In contrast, the ISWT is simple and does not require specialised equipment or extensive training to conduct the tests. The externally-paced and incremental characteristics of the ISWT are also similar to laboratory CPETs,[1] making the ISWT an excellent alternative to assess exercise capacity in clinical settings.

The European Respiratory Society (ERS) and the American Thoracic Society (ATS) have established standardised English instructions for the ISWT.[17] However, its use with non-English speaking populations in regions and countries such as Hong Kong, Taiwan, Malaysia, China, and Singapore may be limited, given the difficulty in understanding the instructions. The Chinese (Mandarin) language is the second-most spoken language worldwide,[18] and approximately 75% of Singaporeans are Chinese, with 48% of them have Chinese as their primary language.[19] This highlights the need for Chinese instructions. However, there is no published validated Chinese version of the ISWT instructions to date. As a result, clinicians have to translate the ISWT instructions on an *ad-hoc* basis, which affects the reliability and validity of the test since translation can differ significantly between individuals, warranting the need for additional psychometric testing on an established measurement after cross-cultural translation and adaptation.[20] This is further attested by the results from the European Social Survey, which showed differences in measurement

quality across countries. Their study attempted to explain such differences and found that mistranslations and/or keywords lost in translations could affect the investigation results.[21] Therefore, this study aimed to develop a Chinese (Mandarin) version of the ISWT instructions (ISWT-CHN) that is conceptually equivalent to the original English version (ISWT-ENG) and subsequently established its reliability and validity.

Methods

Design

This was a translation and cross-sectional validity study assessing the reliability and validity of the ISWT after the adaptation to the Chinese (Mandarin) language. This study took place between June 2020 and April 2021. Ethical approval was obtained from the Institutional Review Board of the university (Number: 2020022), and all subjects provided written informed consent before participating in the study.

Development of ISWT-CHN

Before conceptualising this study, permission was first sought from the test originator.[1] The development of the ISWT-CHN adopted the cross-cultural adaptation process recommended by the World Health Organisation (WHO).[22] Figure 1 describes the translation and cross-cultural adaptation process of the ISWT-CHN. Two bilingual translators performed the forward translations independently. These two versions (Versions A and B) were combined to form the interim instructions (Version C), and it was then compared and revised by two investigators and the two forward translators in agreement to form the revised translated version (Version D). Twenty volunteers who were bilingual in English and Chinese but had no prior experience with the ISWT performed the backward translation of the revised-Chinese version (Version D) to English (Versions E1–E20). Lastly, the panel of researchers, which consisted of four investigators and two translators, reviewed the 20 backward translated versions for face and content validity and modified them to produce the final ISWT-CHN instructions (Version F).

Participants, investigators and centres

The participants were recruited via convenience sampling, with the following inclusion criteria: (1) Between 21 and 65 years old; (2) ambulant and not

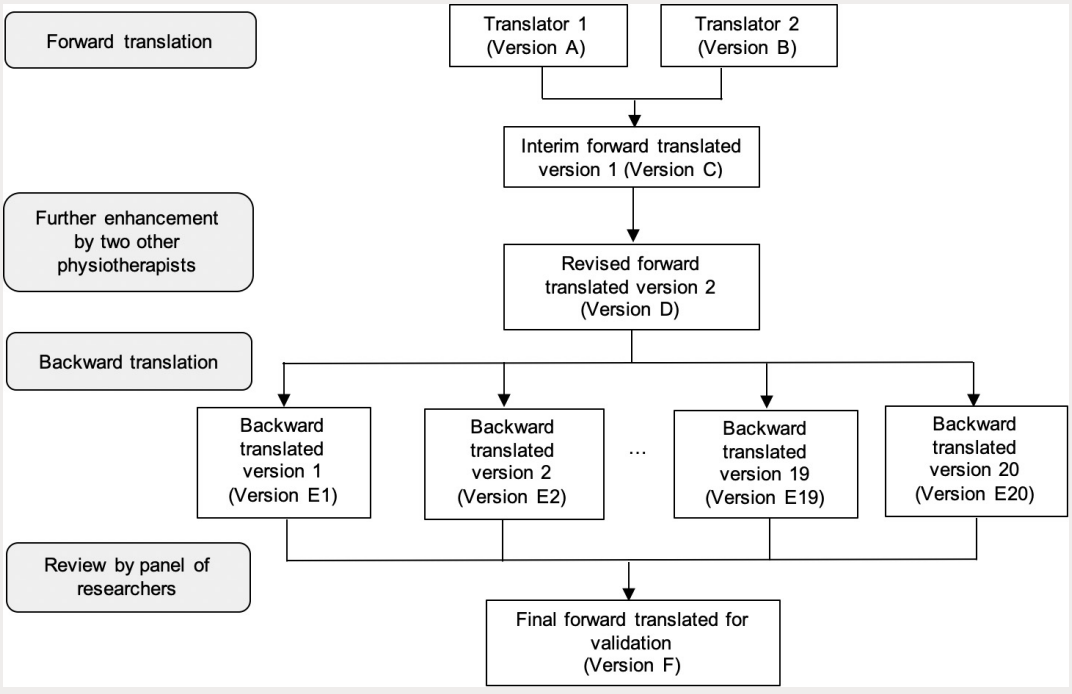


Fig. 1. Flow diagram of the translation and cross-cultural adaptation process.

using any walking aids; (3) colloquially proficient in Chinese and English; (4) has adequate mental capacity and able to follow instructions; and (5) was not involved with the backward translation in the development of the ISWT-CHN. Participants were excluded from the study if they had any cardiovascular, respiratory, neuromuscular and/or musculoskeletal disorder(s) that could hinder their ability to exercise. The Physical Activity Readiness Questionnaire for Everyone (PAR-Q+)[23] was used for screening. Prospective participants were excluded if they were presented with any contraindications for CPET[24] or spirometry results that suggested any potential airflow limitation, defined as a forced expiratory volume in one second (FEV_1) to forced vital capacity (FVC) ratio of less than 0.70.[25,26] According to other cross-cultural adaptation studies[27–31] and the minimum sample size suggested by statisticians,[32] 30 participants would provide adequate study power to observe the validity and reliability. All exercise testing, namely ISWT and CPET, were conducted by two investigators (Rater A and Rater B), undergraduate physiotherapy students with proficiency in conducting the tests, under close supervision from the principal investigator. All testing procedures and data collection took place in the exercise laboratory of the university.

The ISWT protocol

With adherence to the Technical Standard,[17] the ISWT was conducted with an open 10 m course marked by two cones that were placed 0.5 m inwards from either end. Participants were required to listen to a pre-recorded standardised ISWT instruction in either English or Chinese (Mandarin) before the test started and had to keep up with the walking speed dictated by the pre-recorded audio signals while walking up and down the course during the test. The test was terminated if the participant was limited by (1) fatigue; (2) dyspnoea; or (3) inability to maintain the required speed and failed to complete two consecutive shuttles. The distance covered was calculated from the total number of completed shuttles (ISWD). Parameters such as oxygen saturation (SpO_2), heart rate (HR), blood pressure (BP), and the modified Borg's scale (0–10) for Dyspnoea and Rate of Perceived Exertion (RPE)[33,34] were taken at various time points of the test. There was a minimum 30 min break between each test, and the SpO_2, HR, and BP had to return to baseline before the subsequent trial.

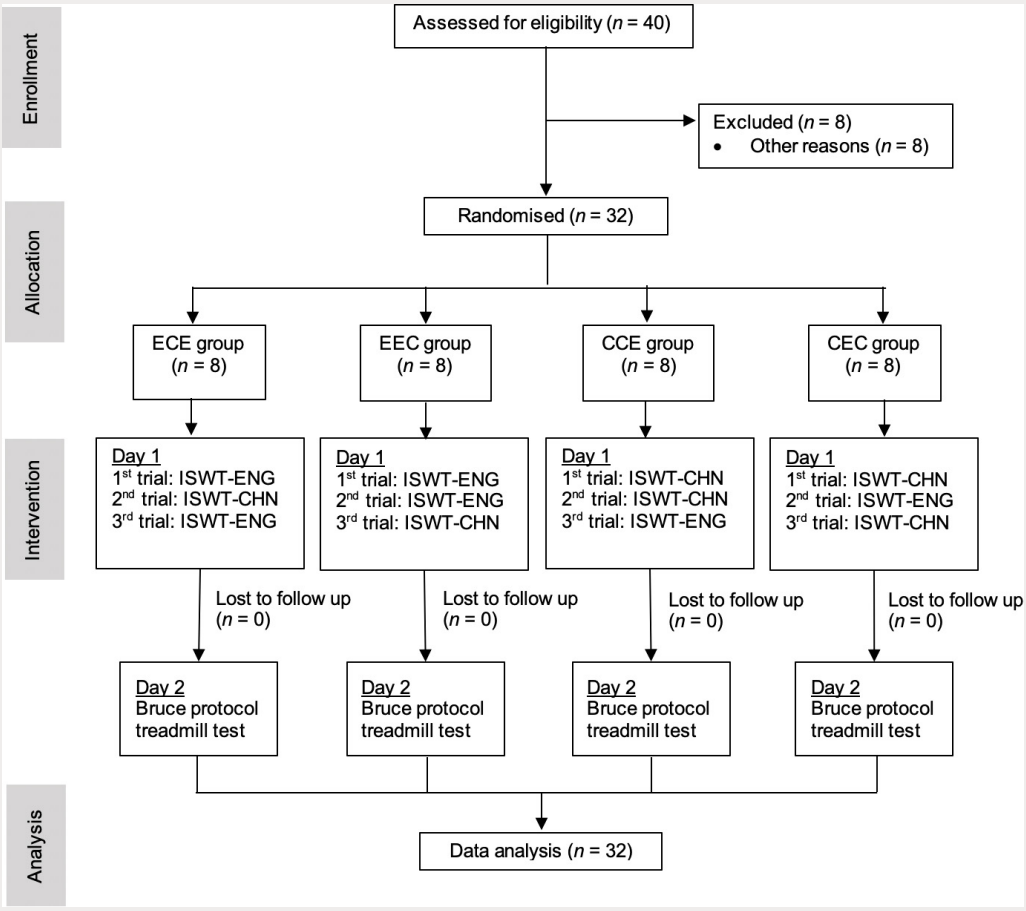
Bruce protocol treadmill test

The CPET was conducted on a treadmill (h/p/cosmos quasar® med, Germany) with metabolic gas analysis (COSMED Quark CPET Gas Analyser, Germany). The CPET was conducted using the standard Bruce treadmill protocol.[35] Participants were instructed to continue the test until maximal exhaustion while trained investigators monitored the necessary parameters [HR and rhythm via a 12-lead electrocardiogram, BP and RPE (0–10)]. The maximal oxygen consumption (VO_2max) recorded by the cardiopulmonary diagnostic software (OMNIA 1.6.3, COSMED, Rome, Italy) was used for data analysis. The CPET results served as the benchmark for criterion validity testing between ISWT-ENG and ISWT-CHN. The ISWT and CPET testings' were conducted on separate days.

Reliability and validity of ISWT-CHN

A pilot trial was first conducted on six healthy participants who fulfilled the same aforementioned inclusion and exclusion criteria with the ISWT-CHN instructions (Version F) to finalise the translation and establish the inter-reliability of the ISWT-CHN (Raters A and B). Both investigators contributed to the collection of the demographic data. Each participant completed a total of five trials on the same day. Considering the learning effect on the ISWT,[36,37] a practice trial was incorporated before the participants were randomised into A-B-B-A ($n = 3$) or B-A-A-B ($n = 3$) sequence via a computer-generated list for optimal efficiency and inter-rater reliability estimates.[38] The better of the two ISWD from each rater was used to establish the inter-rater reliability of the ISWT-CHN. No further changes to the ISWT-CHN instructions were made after the pilot trial, and Version F was concluded as the final ISWT-CHN instructions. Subsequently, the intra-rater reliability was established with an additional 14 healthy participants.

In the validity study, participants performed three trials of ISWT on the same day and the CPET on the following day. The block randomisation method with a block size of three was used. The block size of three was the minimum



Notes: ECE = English–Chinese–English, EEC = English–English–Chinese, CEC = Chinese–English–Chinese, CCE = Chinese–Chinese–English, ISWT-CHN = Chinese (Mandarin) version of the Incremental Shuttle Walk Test, ISWT-ENG = original version of the Incremental Shuttle Walk Test.

Fig. 2. Design and flow of participants through the validity study.

permutation for two variables, yet it was able to reduce bias and achieve balance in the allocation of participants compared to simple randomisation, particularly applicable with a small sample size.[39] The recruited participants were randomised using a computer-generated list into four groups with different test sequences for the validity testing of the ISWT-CHN. The groups were as follows: The English–Chinese–English (ECE; $n = 8$), English–English–Chinese (EEC; $n = 8$), Chinese–English–Chinese (CEC; $n = 8$) or Chinese–Chinese–English (CCE; $n = 8$) groups (Fig. 2). The better result of the two trials in the same language and the result of the remaining trial in the other language within the same test sequence, were chosen for analysis to account for the learning effect.[36,37] For example, the better result of the two English ISWD and the Chinese ISWD from the ECE group were used for subsequent analysis.

Data analysis

Statistical analysis was performed with IBM SPSS® Statistics for Windows Version 26.0 (IBM Corporation, Armonk, New York, USA). The level of significance was set at $p < 0.05$. Demographic and anthropometric data of participants were examined for normal distribution using the Shapiro–Wilk test. ANOVA was used to analyse normally distributed continuous variables between groups (Height, Weight, FEV_1/FVC ratio, Chinese and English ISWD, Absolute $VO_{2\,max}$) and the Kruskal–Wallis test was used to analyse non-normal continuous variables (Age). The Fisher's exact test was used to analyse sex differences between groups.

The inter-rater and intra-rater reliability were analysed via the interclass correlation (ICC) with a two-way mixed-effects model and a Bland–Altman plot.[40] The ICC was based on a 95% Confidence Interval (95% CI). Absolute reliability

was calculated with the standard error of measurement (SEM), minimal detectable change (MDC), and 95% limits of agreement (LOA). SEM was calculated based on the formula SD $\sqrt{(1 - \text{ICC})}$, while MDC at 95% level of confidence was calculated based on the formula $\text{SEM} \times \sqrt{2} \times 1.96$. The 95% LOA provides an interval within which 95% of differences between the two measurements are expected to lie. The alternate form of reliability between the two measurements can be ensured if the 95% LOA is narrow and at least 95% of the points lie within the limits.[41,42]

Face validity was determined during the translation process. Content validity was determined by the CVI and CVR.

(1) CVI is the most widely reported method for determining content validity in instrument development that examines its relevance and clarity, and it can be calculated using the Item-CVI (I-CVI) and the Scale-level-CVI (S-CVI).[43] This study used a 4-point Likert scale (not at all, needs some revision, needs minor revision, complete), where the I-CVI was calculated by the total ratings scored by all the panel members ($n = 6$; four investigators and two translators) and divided by the total number of panel members. Where I-CVI is greater than 0.79, the item is relevant; between 0.70 and 0.79, the item requires revisions; and when it is less than 0.70, the item is eliminated.[43,44] Similarly, S-CVI is determined by the number of items in the instrument that received a "highly complete" grade. The Universal Agreement (UA) among the panel members (S-CVI/UA) and the Average CVI (S-CVI/Ave) are two ways of determining S-CVI, the latter being a less conservative method.[43] S-CVI/UA is calculated by the sum of all items with I-CVI equal to 1 divided by the total number of items, whereas S-CVI/Ave is equal to the sum of all the I-CVIs divided by the number of items. Content validity is excellent when the S-CVI/UA is more than 0.8 and the S-CVI/Ave is more than 0.9.[44] This study used the S-CVI/UA method.
(2) CVR measures the essentiality of an item[45] CVR ranges from −1 to 1, with a higher score indicating greater agreement among panel members. $\text{CVR} = (\text{Ne} - \text{N}/2)/(\text{N}/2)$, where Ne is the number of panellists who rated an item as "essential" and N is the total number of panellists.[43] Each sentence's essentiality was examined on a 3-point Likert scale (not essential, useful but not essential, essential). For a panel size of six, $\text{CVR} = 1.0$ ($p = 0.05$)[46] is required to be statistically significant.

Construct and criterion validities were analysed using Pearson's correlation coefficient r. Construct validity was determined from the correlation between the distances obtained from ISWT-ENG and ISWT-CHN, while criterion validity was determined from the correlations between the ISWDs of the ISWT-CHN, ISWT-ENG and $VO_{2\,max}$ of the CPET.

Results

Flow of participants

Six participants were recruited during the pilot trial to establish the inter-rater reliability, and an additional 14 participants were included to establish the intra-rater reliability. All screened participants fulfilled the inclusion criteria, and none were excluded according to the exclusion criteria. Table 1 presents the overall characteristics of the participants in the reliability study. Subsequently, 40 participants were assessed for eligibility during the validity study, and eight participants were excluded according to the exclusion criteria. The remaining 32 participants, 20 males (63%) and 12 females (37%) satisfied the inclusion and exclusion criteria and participated in the validity study. The demographic and clinical data (age, sex, height, weight, spirometry readings, ISWD, and $VO_{2\,max}$)

Table 1. Demographics of participants in the reliability study ($n = 20$).

	Participants
Age (years)	26.7 (SD 5.1)
Sex	
Male	11
Female	9
Height (cm)	169 (SD 7.6)
Weight (kg)	67 (SD 14.9)
FEV_1/FVC ratio (%)	86 (SD 5.0)

Notes: SD: Standard deviation; cm: centimetres; kg: kilograms; FEV_1/FVC ratio (%): forced expiratory volume in one second (FEV_1) to forced vital capacity (FVC) ratio in percentage.

Table 2. Demographics and clinical data of participants in the validity study ($n = 32$).

	ECE ($n = 8$)	EEC ($n = 8$)	CEC ($n = 8$)	CCE ($n = 8$)	p-value
Age (years)	24 (SD 2.2)	27 (SD 3.7)	27 (SD 6.5)	24 (SD 3.1)	0.23
Sex					0.96
Male ($n = 20$)	6	5	5	4	
Female ($n = 12$)	2	3	3	4	
Height (cm)	168 (SD 5.4)	165 (SD 6.0)	168 (SD 8.1)	170 (SD 7.3)	0.53
Weight (kg)	62 (SD 7.3)	60 (SD 9.2)	72 (SD 16.0)	62 (SD 12.0)	0.18
FEV_1/FVC ratio (%)	90 (SD 3.3)	85 (SD 6.0)	85 (SD 3.0)	88 (SD 6.5)	0.15
Chinese ISWD (m)	699 (SD 155.0)	801 (SD 181.0)	738 (SD 150.0)	803 (SD 163.0)	0.52
English ISWD (m)	716 (SD 136.0)	801 (SD 174.0)	713 (SD 160.0)	806 (SD 157.0)	0.47
Absolute $VO_{2\,max}$ (mL/min/kg)	47 (SD 8.7)	40 (SD 8.1)	41 (SD 14.0)	47 (SD 16.0)	0.52

Notes: ECE = English–Chinese–English, EEC = English–English–Chinese, CEC = Chinese–English–Chinese, CCE = Chinese–Chinese–English, n = number of participants, ISWD = Incremental Shuttle Walk Distance, FEV_1 = forced expiratory volume in one second, FVC = forced vital capacity, and $VO_{2\,max}$ = maximal oxygen consumption. *Significance of p values < 0.05.

of the participants in the validity study are summarised in Table 2.

Reliability of ISWT-CHN

Table 3 illustrates the inter- and intra-rater reliability. The ISWT-CHN displayed excellent inter-rater reliability (ICC = 0.99, 95% CI 0.97–1.0, $p < 0.001$) between Raters A and B, with the SEM and MDC being 0.85 metres (m) and 2.35 m, respectively. Figure 3 shows the Bland–Altman plot for the inter-rater reliability, and the dots on the plot represents each of the six participants for the pilot study. From the plot, all the points were within the 95% LOA (+95% LOA = 37 m, −95% LOA = −37 m). Good intra-rater reliability (Rater A: ICC = 0.92, 95% CI 0.53–0.98, $p = 0.003$; Rater B: ICC = 0.90, 95% CI 0.76–0.96, $p < 0.001$) was observed. The SEM values for the intra-rater reliability testing were 35 m for Rater A and 32 m for Rater B. The MDC values for the intra-rater reliability testing were 97 m for Rater A and 88 m for Rater B. The SEM values for the reliability testing could be considered small as they were below 10% of the highest achievable distance of 1020 m.[47]

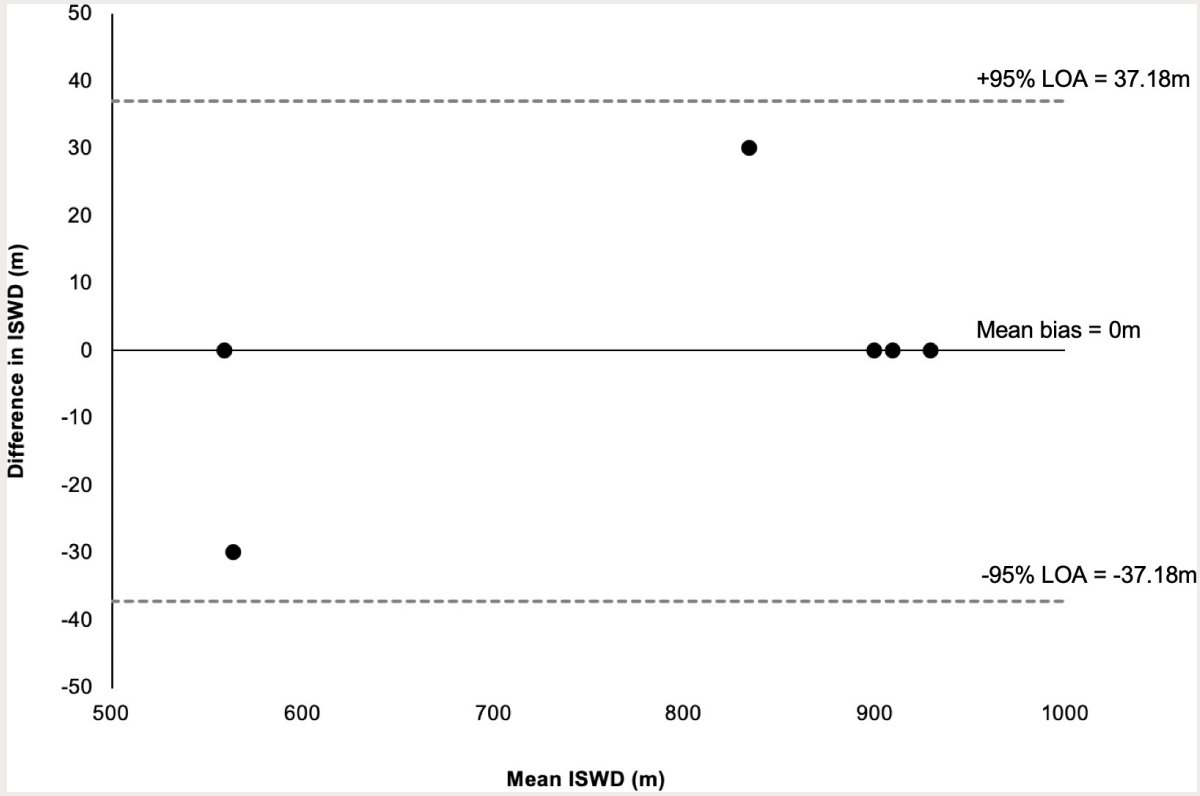
Validity of ISWT-CHN

The rigorous translation process established the face validity of the ISWT-CHN according to the translation and cross-cultural adaptation guidelines by the WHO.[22] The panel of six investigators synonymously agreed that the face meaning of the 20 copies of backward translations collected was equivalent to the original English instructions. Subsequently, content validity was established via quantitative and qualitative assessments.

Table 3. Reliability of the ISWT-CHN.

	ICC	95% CI	p-values	SEM (m)	MDC (m)
Inter-rater reliability ($n = 6$)	0.99	0.97-1.00	< 0.001*	0.85	2.35
Intra-rater reliability ($n = 20$)					
Rater A	0.92	0.53-0.98	$= 0.003$*	35.0	97.0
Rater B	0.90	0.76-0.96	< 0.001*	32.0	88.0

Notes: ISWT-CHN = Chinese (Mandarin) version of the Incremental Shuttle Walk Test, ICC = Intra-class Correlation Coefficient, CI = Confidence Interval, SEM = Standard Error of Measurement, and MDC = Minimal Detectable Change. *Significance of p values < 0.05.



Notes: The differences in incremental shuttle walk distances between raters are plotted against the mean scores. The straight line represents the mean difference between the two raters; dashed lines represent the 95% limits of agreement. ISWD = Incremental Shuttle Walk Distances and LOA = limits of agreement.

Fig. 3. Bland–Altman plot comparing the agreement between two raters.

Six investigators provided feedback and ratings on the essentiality, relevance, and clarity of each sentence of the translated instruction. There was unanimous agreement that the conceptual meaning of the IWST-CHN was consistent with the original version. The I-CVI, S-CVI/Ave and CVR of the ISWT-CHN were 1.0.

The ISWT-CHN had a mean distance of 760 m [standard deviation (SD) 161] and ISWT-ENG had a mean distance of 759 m (SD 156). The Pearson's coefficient of ISWT-CHN versus CPET ($r = 0.439$, $p < 0.001$) as compared to ISWT-ENG versus CPET ($r = 0.448$, $p < 0.001$). This similarly weak positive correlation with the $VO_{2\,max}$ established the criterion validity of the ISWT-CHN. Additionally, the ISWT-CHN had satisfactory construct validity, as shown by the very high positive correlation between the ISWD of the ISWT-CHN and ISWT-ENG ($r = 0.967$, $p < 0.001$). Figure 4 shows the scatterplots that depict the relationships between $VO_{2\,max}$ (mL/min/kg), ISWT-CHN distance (m) and ISWT-ENG distance (m).

Discussion

This study aimed to cross-culturally adapt the English ISWT instructions to Chinese (Mandarin) and evaluate the translation's reliability and validity. This study revealed that ISWT-CHN is a reliable and valid outcome measure. Intra-rater reliability of both Raters A and B were good, with an ICC of 0.92, $p = 0.003$ and 0.90, $p < 0.001$, respectively. Inter-rater reliability was excellent with an ICC of 0.99, $p < 0.001$, and the Bland–Altman plot revealed all points within ±1.96 SD of the mean difference implying that the two raters agreed. Absolute reliability measurements also showed that the ISWT-CHN results were precise and had low margins of error between raters, with an SEM of 0.85 m and an MDC of 2.35 m. This agrees with the results from other studies that showed good to excellent reliability, for instance, in the COPD population (ICC = 0.80–0.93)[2] and patients with peripheral obstructive arterial disease (ICC = 0.95).[9]

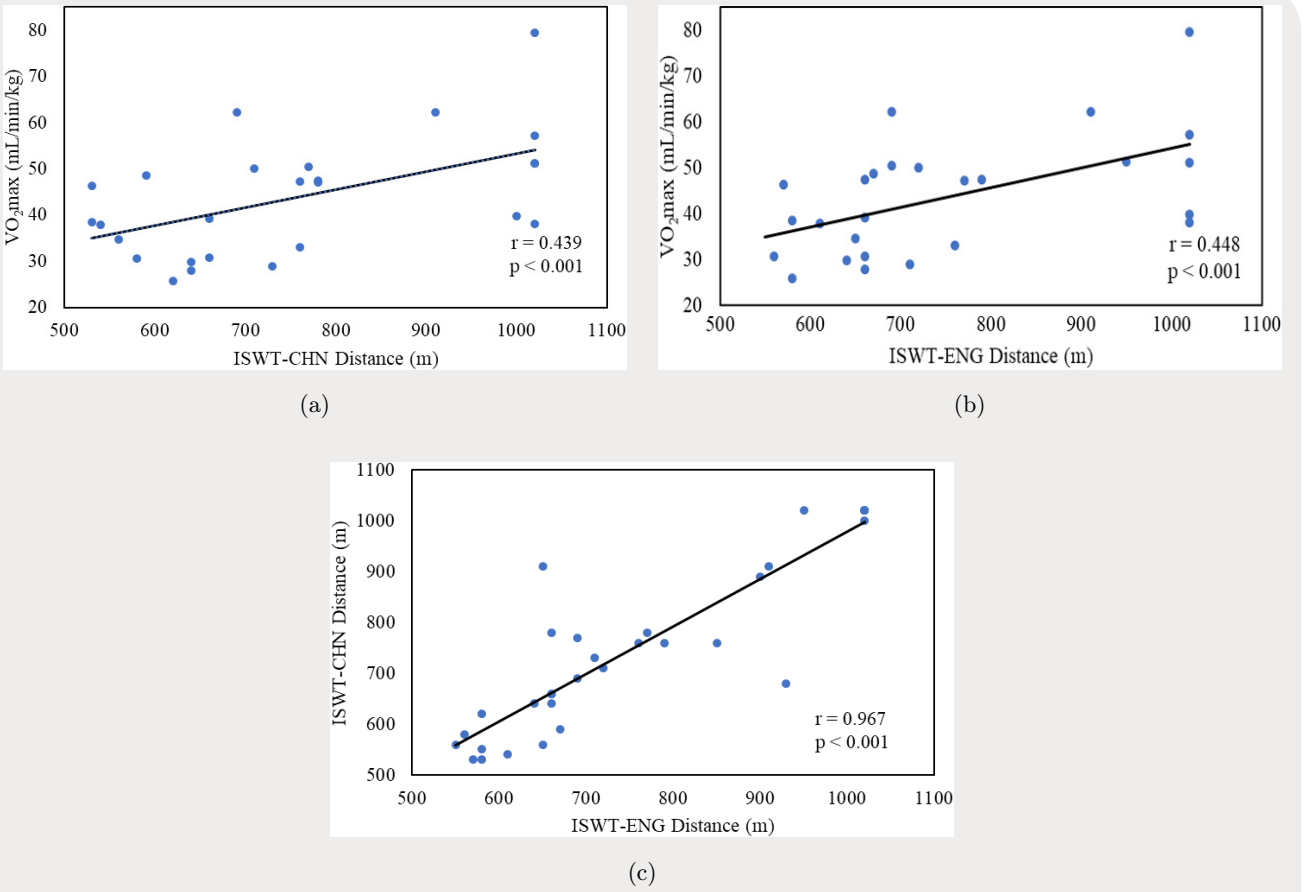


Fig. 4. Scatterplots of $VO_{2\,max}$, ISWT-CHN distance and ISWT-ENG distance. (a) $VO_{2\,max}$ and ISWT-CHN distance, (b) $VO_{2\,max}$ and ISWT-ENG distance, (c) ISWT-CHN distance and ISWT-ENG distance. ISWT-CHN = Chinese (Mandarin) version of the Incremental Shuttle Walk Test, ISWT-ENG = original version of the Incremental Shuttle Walk Test, $VO_{2\,max}$ = maximal oxygen consumption, r = Pearson's correlation, and p = probability value.

The validity of the ISWT-CHN was established. Face validity was determined during the translation process. Content validity was ascertained by the determination of CVI, in the form of I-CVI and S-CVI, as well as CVR. Construct and criterion validity of ISWT-CHN were also established. The construct validity of the ISWT-CHN was established through the excellent correlation with the ISWT-ENG ($r = 0.967$, $p < 0.001$). The ISWT-ENG and ISWT-CHN results were compared with CPET $VO_{2\,max}$ results, the gold standard to measure exercise capacity.[16] The criterion validity of the ISWT-CHN was indicated by the largely similar significant positive correlation of the ISWT-CHN and ISWT-ENG with the $VO_{2\,max}$($r = 0.439, p < 0.001; r = 0.448, p < 0.001$), which suggested strong agreement between both versions of ISWT. The similar degree of weak positive correlation between the $VO_{2\,max}$ and ISWT in both languages and the strong correlation between the ISWT-CHN and ISWT-ENG demonstrated the equivalence of the instructions in both the languages. Therefore, the establishment of criterion validity was inferred.

This study revealed a weaker correlation between ISWT results in both languages and CPET $VO_{2\,max}$ results ($r = 0.439$–0.448) as compared to other studies. The Pearson's correlation coefficient r ranged from 0.72 to 0.85 in the COPD population,[3–5] from 0.73 to 0.83 in patients with cardiac diseases,[6,7] and was 0.67 in lung cancer patients.[8] A possible reason for this is because the original ISWT[1] was developed for patients with COPD, and the protocol does not allow running to keep up with the set speeds. This, however, might not be sensitive to predict functional exercise capacity in the healthy population as they are less likely to reach 85% of the predicted maximal HR or become too breathless to continue just by walking. Moreover, when observing the best ISWT trial results

regardless of the language used in this study, none of the participants terminated the test due to excessive dyspnoea and/or leg fatigue.[1] In fact, most participants (78%, $n = 25$) terminated the test due to the inability to maintain the required speed.[1] Hence, in most participants, functional exercise capacity could not be accurately predicted as the test was terminated before adequate cardiovascular stress. The remaining 22% of participants ($n = 7$) completed all 12 levels at least once out of the three ISWT trials, and 86% ($n = 6$) of them were males. Functional exercise capacity could not be accurately predicted in this subgroup of participants due to ceiling effects. To overcome these limitations when using ISWT in a healthy population, Probst *et al.*[13] suggested allowing the participants to run if needed and increasing the number of stages beyond 12 levels by increasing the speed by 0.17 metres per second (m/s) each minute. However, this modified protocol[37] is not commonly used in clinical settings.

The ISWT is widely used in research and rehabilitation as a measure of exercise capacity. It is often chosen over other measures of exercise capacity, such as the 6-min walk test (6MWT), as this test necessitates at least 30 m of walking space[48] which may not be available at clinical or research facilities. Furthermore, given that the 6MWT is associated with several limitations related to its self-paced nature, the ISWT is superior when determining exercise intensity because of its progressive nature.[48,49] The ISWT is relevant and particularly useful when prescribing exercise intensity using the percentage of peak performance[48] and thus may be more beneficial in the rehabilitation or general health promotion setting when tailoring individualised exercise intervention.[11–13] The newly translated Chinese instructions of the ISWT may offer a reliable and validated method of field-based exercise testing, benefiting Chinese-speaking populations around the globe.

There are several limitations of this study. Firstly, the study used convenience sampling and recruited only a small sample size, which might have introduced bias where specific subgroups of populations might have been over- or under-represented. For instance, our participants were relatively young, with a mean age of 26.7 (SD 5.1) and 25 (SD 4.2) years in the reliability and validity studies, respectively, although the study's age inclusion criteria were set between 21 and 65 years old. This also resulted in a non-normal distribution with data skewed to the left. This may be due to the study's recruitment process, which required the participants to be on social media to obtain information about the study. This method may have limited the study's reach to only those proficient in technology navigating the social media platforms. Furthermore, a portion of interested participants who were in older age groups chose not to participate in the study after understanding what CPET was, and declined with the main reason being "afraid of over-exerting". Although interested and willing to perform CPET, some older adults were excluded from the study as they had chronic medical conditions and were deemed unsafe for exercise until certified by qualified personnel, as stated in the PAR-Q$^+$. Additionally, it is worth noting that the study was conducted amid the COVID-19 global pandemic, where multiple measures have been put in place, such as movement restrictions and proper social distancing to curb transmissions. Thus, participant recruitment was mainly limited to the younger population, and the sample size fell short of the 30 participants required to provide adequate study power. Notwithstanding, the reliability results are promising despite the small sample size. It is also unclear if there exists a difference in colloquial usage and understanding of the language between age groups which warrants further exploration.

Nonetheless, this study is the first to develop a Chinese version of the ISWT instructions with established reliability and validity. The study results were comparable to other published data, and it showed significant results to confirm that the ISWT-CHN was conceptually equivalent to its original English instructions. Thus, this study can serve as a framework for future studies with larger sample sizes and participants of varied backgrounds despite its limitations.

Conclusion

This study translated and cross-culturally adapted the original English instructions of the ISWT into the Chinese (Mandarin) language and subsequently established its reliability and validity. The ISWT-CHN was shown to be a reliable and valid measurement of exercise capacity with psychometric properties similar to the original English version.[1] This development is helpful to the Chinese-speaking population as clinicians can now

conduct the ISWT using standardised Chinese instructions instead of performing *ad-hoc* translation. However, given the limited sample size, future research on the ISWT-CHN with a larger sample size and participants of different backgrounds is recommended.

Acknowledgment

We are grateful to the participants who took their time off to participate in this study.

Conflict of Interest

The author(s) declared no potential conflicts of interest concerning this paper's research, authorship, and/or publication.

Funding/Support

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this paper. This work was supported by the Singapore Institute of Technology, Grant No. R-MOE-A404-H009.

Ethics Approval

The Singapore Institute of Technology-Institutional Review Board (SIT-IRB Project Number: 2020022) approved this study. All participants gave written informed consent before data collection began.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

Wei Qin Ang and Hong Ting Tan collected, analysed and interpreted the data, prepared and revised the paper critically for important intellectual content. Si Min Goh collected, analysed, and interpreted the data. Samantha W. Seng collected and interpreted the data. Katherin S. Huang and Melissa Y. Chan conceptualised and designed the study, drafted and revised the paper critically for important intellectual content. Meredith T. Yeung conceptualised and designed the study, analysed and interpreted the data, revised the paper critically for important intellectual content and approved the final paper. Wei Qin Ang and Hong Ting Tan are the co-first authors who contributed and credited equally to this work.

References

1. Singh SJ, Morgan MD, Scott S, Walters D, Hardman AE. Development of a shuttle walking test of disability in patients with chronic airways obstruction. Thorax 1992;47:1019–24.
2. Campo LA, Chilingaryan G, Berg K, Paradis B, Mazer B. Validity and reliability of the modified shuttle walk test in patients with chronic obstructive pulmonary disease. Arch Phys Med Rehabil 2006;87:918–22.
3. Onorati P, Antonucci R, Valli G, et al. Non-invasive evaluation of gas exchange during a shuttle walkingest vs. a 6-min walking test to assess exercise tolerance in COPD patients. Eur J Appl Physiol 2003;89:331–6.
4. Satake M, Shioya T, Takahashi H, Kawatani M. Ventilatory responses to six-minute walk test, incremental shuttle walking test, and cycle ergometer test in patients with chronic obstructive pulmonary disease. Biomed Res 2003;24: 309–16.
5. Turner SE, Eastwood PR, Cecins NM, Hillman DR, Jenkins SC. Physiologic responses to incremental and self-paced exercise in COPD: A comparison of three tests. Chest 2004;126:766–73.
6. Green DJ, Watts K, Rankin S, Wong P, O'Driscoll JG. A comparison of the shuttle and 6 minute walking tests with measured peak oxygen consumption in patients with heart failure. J Sci Med Sport 2001;4:292–300.
7. Lewis ME, Newall C, Townend JN, Hill SL, Bonser RS. Incremental shuttle walk test in the assessment of patients for heart transplantation. Heart 2001;86:183–7.
8. Win T, Jackson A, Groves AM, Sharples LD, Charman SC, Laroche CM. Comparison of shuttle walk with measured peak oxygen consumption in patients with operable lung cancer. Thorax 2006;61:57–60.
9. da Cunha-Filho IT, Pereira DA, de Carvalho AM, Campedeli L, Soares M, de Sousa Freitas J. The reliability of walking tests in people with claudication. Am J Phys Med Rehabil 2007;86:574–82.
10. Arena R, Myers J, Williams MA, et al. Assessment of functional capacity in clinical and research settings. Circulation 2007;116:329–43.

11. Harrison SL, Greening NJ, Houchen-Wolloff L, et al. Age-specific normal values for the incremental shuttle walk test in a healthy British population. J Cardiopulm Rehabil Prev 2013;33:309–13.
12. Itaki M, Kozu R, Tanaka K, et al. Reference equation for the incremental shuttle walk test in Japanese adults. Respir Investig 2018;56:497–02.
13. Probst VS, Hernandes NA, Teixeira DC, et al. Reference values for the incremental shuttle walking test. Respir Med 2012;106:243–8.
14. Emtner MI, Arnardottir HR, Hallin R, Lindberg E, Janson C. Walking distance is a predictor of exacerbations in patients with chronic obstructive pulmonary disease. Respir Med 2007;101:1037–40.
15. Nutt CL, Russell JC. Use of the pre-operative shuttle walk test to predict morbidity and mortality after elective major colorectal surgery. Anaesthesia 2012;67:839–49.
16. Ferrazza AM, Martolini D, Valli G, Palange P. Cardiopulmonary exercise testing in the functional and prognostic evaluation of patients with pulmonary diseases. Respiration 2009;77:3–17.
17. Holland AE, Spruit MA, Singh SJ. How to carry out a field walking test in chronic respiratory disease. Breathe (Sheffield, England) 2015; 11:128–39.
18. Statista. The most spoken languages worldwide in 2021. https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/: Statista, 2021.
19. DoS. Singapore Census of Population 2010, Statistical Release 1. Demographic Characteristics, Education, Language and Religion. Department of Statistics, 2011.
20. Peña ED. Lost in translation: Methodological considerations in cross-cultural research. Child Dev 2007;78:1255–64.
21. Oberski D, Saris WE and Hagenaars J. Why are there differences in measurement quality across countries. In: Loosveldt G, Swyngedouw M and Billiet J, (eds.). Measuring Meaningful Data in Social Research, 1st ed. Leuven: Acco, 2007.
22. WHO. Process of translation and adaptation of instruments. Wold Health Organization, 2020.
23. Warburton DE, Jamnik VK, Bredin SS, Gledhill N. The physical activity readiness questionnaire for everyone (PAR-Q+) and electronic physical activity readiness medical examination (ePARmed-X+). Health Fit J Canada 2011;4:3–17.
24. ATS/ACCP Statement on cardiopulmonary exercise testing. Am J Respir Crit Care Med 2003;167:211–77.
25. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 2007;176:532–55.
26. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. Eur Respir J 2005;26:319–38.
27. Macfarlane DJ, Lee CC, Ho EY, Chan KL, Chan DT. Reliability and validity of the Chinese version of IPAQ (short, last 7 days). J Sci Med Sport 2007;10: 45–51.
28. Poon DM, Chow LC, Au DK, Hui Y, Leung MC. Translation of the dizziness handicap inventory into Chinese, validation of it, and evaluation of the quality of life of patients with chronic dizziness. Ann Otol Rhinol Laryngol 2004;113:1006–11.
29. Dhippayom JP, Trevittaya P, Cheng ASK. Cross-cultural adaptation, validity, and reliability of the patient-rated Michigan hand outcomes questionnaire for Thai patients. Occup Ther Int 2018;2018:8319875.
30. Passos L, Tavares J, Figueiredo D. Cross-cultural adaptation and analysis of the psychometric properties of the Portuguese version of the Edinburgh feeding evaluation in Dementia scale. Port J Public Health 2020;38:142–50.
31. Leung RW, Leung M-L, Chung P-K. Validity and reliability of a Cantonese-translated rating of perceived exertion scale among Hong Kong adults. Percept Mot Ski 2004;98:725–35.
32. Onwuegbuzie AJ, Collins KM. A typology of mixed methods sampling designs in social science research. Qual Rep 2007;12:281–16.
33. Borg GA. Perceived exertion. Exerc Sport Sci Rev 1974;2:131–53.
34. Borg GA. Psychophysical bases of perceived exertion. Med Sci Sports Exerc. 1982;14:377–81.
35. Bruce RA, Blackmon JR, Jones JW, Strait G. Exercising testing in adult normal subjects and cardiac patients. Pediatrics 1963;32:742–56.
36. Dyer F, Marriner P, Cheema K, Bott J. Is a practice incremental shuttle walk test really necessary? Chronic Respir Dis 2011;8:201–5.
37. Bradley J, Howard J, Wallace E, Elborn S. Validity of a modified shuttle test in adult cystic fibrosis. Thorax 1999;54:437–9.
38. Hallgren KA. Computing inter-rater reliability for observational data: An overview and tutorial. Tutor Quant Meth Psychol. 2012;8:23–34.
39. Efird J. Blocked randomization with randomly selected block sizes. Int J Environ Res Public Health. 2010;8:15–20.
40. Bland JM, Altman DG. Measurement error. Bmj 1996;313:744.
41. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med. 1998;26:217–38.
42. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000;30:1–15.

43. Zamanzadeh V, Ghahramanian A, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar A-R. Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. J Caring Sci 2015;4:165–78.
44. Polit DF, Beck CT. The content validity index: Are you sure you know what's being reported? critique and recommendations. Res Nursing Health 2006;29:489–97.
45. Yamada J, Stevens B, Sidani S, Watt-Watson J, De Silva N. Content validity of a process evaluation checklist to measure intervention implementation fidelity of the EPIC intervention. Worldviews Evid-Based Nurs 2010;7: 158–64.
46. Ayre C, Scally AJ. Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. Meas Eval Couns Dev 2013;47:79–86.
47. Prince B, Makrides L, Richman J. Research methodology and applied statistics. Part 2: The literature search. Physiother Can 1980;32:201–6.
48. Holland AE, Spruit MA, Troosters T, et al. An official European Respiratory Society/American Thoracic Society technical standard: Field walking tests in chronic respiratory disease. Eur Respir J 2014;44:1428–46.
49. Wise RA, Brown CD. Minimal clinically important differences in the six-minute walk test and the incremental shuttle walking test. J Chron Obstruct Pulmon Dis 2005;2:125–9.