Validity and reliability of a new incremental step test for people with chronic obstructive pulmonary disease

Rui Vilarinho,1,2 Lúcia Serra,2 Ana Águas,2 Carlos Alves,3,4 Pedro Matos Silva,1,5 Cátia Caneiras,2,6,7 António Mesquita Montes1,8

ABSTRACT

Background Incremental step tests (IST) can be used to assess exercise capacity in people with chronic obstructive pulmonary disease (COPD). The development of a new step test based on the characteristics of the incremental shuttle walk test (ISWT) is an important study to explore. We aimed to develop a new IST based on the ISWT in people with COPD, and assess its validity (construct validity) and reliability, according to Consensus-based Standards for the selection of health status Measurement Instruments (COSMIN) recommendations.

Methods A cross-sectional study was conducted in participants recruited from hospitals/clinics. During the recruitment, the participants who presented a 6-minute walk test (6MWT) report in the previous month were also identified and the respective data was collected. Subsequently, participants attended two sessions at their homes. IST was conducted on the first visit, along with the 1 min sit-to-stand (1MSTS) test. IST was repeated on a second visit, performed 5–7 days after the first one. Spearman’s correlations were used for construct validity, by comparing the IST with the 6MWT and the 1MSTS. Intraclass correlation coefficient (ICC2,1), SE of measurement (SEM) and minimal detectable change (MDC95) were used for reliability. The learning effect was explored with the Wilcoxon signed-rank test.

Results 50 participants (70.8±7.5 years) were enrolled. IST was significant and moderate correlated with the 6MWT (ρ=0.50, p=0.020), and with the 1MSTS (ρ=0.46, p=0.001). IST presented an ICC2,1=0.96, SEM=10.1 (16.6%) and MDC95=27.9 (45.8%) for the number of steps. There was a statistically significant difference between the two attempts of the IST (p=0.030).

Conclusion Despite the significant and moderate correlations with the 6MWT and 1MSTS, the inability to full compliance with the COSMIN recommendations does not yet allow the IST to be considered valid in people with COPD. On the other hand, the IST is a reliable test based on its high ICC, but a learning effect and an 'indeterminate' measurement error were shown.

Trial registration number NCT04715659.

INTRODUCTION

Improving exercise capacity in people with chronic obstructive pulmonary disease (COPD) is a priority throughout pulmonary rehabilitation (PR) programmes.1–5 PR is a safe, comprehensive, and evidence-based approach in COPD,6 and it can be conducted in a range of settings, where programmes outside the hospital must be implemented, especially home-based programmes.7

In home-based programmes, the use of options to assess exercise capacity with minimal physical space required is more appropriate than the field walking tests normally used: the 6 min walk test (6MWT) and the incremental/endurance shuttle walk test (ISWT/ESWT).8 9 Step tests are a suitable alternative since, in addition to the advantage mentioned above, they require little equipment (an easily transportable platform) and the stepping skill requires little practice.10

According to a systematic review with the aim of identifying the step tests applied to people with COPD,11 nine step tests were identified where most of them present a self-paced work rate profile. Another important observation was that these self-paced step
tests (eg, 6 min step test and 6 min stepper test) were mainly adapted from the 6MWT, also with a self-paced profile, which presents strong measurement properties in people with COPD. However, a field test with an incremental and externally paced profile can have advantages over self-paced tests by providing a symptom-limited maximum response in people with COPD. As an example, the ISWT, originally developed for people with COPD and with strong measurement properties, has advantages over the 6MWT, because it causes an incremental increase in oxygen uptake (VO_{2}), offering an incremental protocol similar to a cardiopulmonary exercise testing (CPET). This facilitates prescribing an exercise regimen as a percentage of peak performance on a field test. According to the systematic review mentioned before, a step test based on the characteristics of the ISWT is not yet available. In fact, only two step tests with an incremental and externally profile were identified, namely the Chester step test (CST) and the Modified Incremental Step Test (MIST). MIST was developed as a modified version of the CST, as the CST was originally developed for healthy subjects, and has important disadvantages when applied in people with COPD, where their performances have a very short duration. Therefore, the development of a step test based on the characteristics of the ISWT is an important area to explore, thus providing an alternative option to assess exercise capacity in people with COPD.

As mentioned before, the ISWT provides a similar response to CPET (same VO_{2peak}) in people with COPD, supporting the choice of the ISWT as an alternative peak test in COPD. In other words, it allows the prescription of endurance training based on the highest speed achieved according to the last level completed on the ISWT, which represents an easy and feasible option in clinical practice. Thus, the same rationale can be applied to a step test by adapting the same number of levels and duration of each level (increment) of the ISWT, in order to collect the highest step cadence achieved according to the last level completed.

The development of a new field test requires the study of their measurement properties, namely its validity and reliability, before its full implementation in clinical practice, to assure that its selection is evidence based. Therefore, this study aimed to develop a new incremental and externally paced step test (IST), and assess its validity (construct validity) in people with COPD. Another aim is to determine its between-days test-retest reliability.

**MATERIAL AND METHODS**

**Study design and sample size**

A cross-sectional study was conducted between March 2020 and July 2021.

For this study, the methodology and sample size were defined following Consensus-based Standards for the selection of health status Measurement Instruments (COSMIN) guidelines. Moreover, validity and reliability were also defined according to COSMIN recommendations. Validity is commonly defined as the extent to which the test can measure the concept it was designed to measure, that is, if it relates to the gold-standard measure (criterion validity) or other measures that assess the same construct (hypotheses testing for construct validity). Based on hypotheses testing for construct validity, a comparison with another outcome measurement instrument was assessed (convergent validity) in this study, by analysing the correlation between the number of steps taken in the IST and the 6 min walk distance (6MWD), and the number of repetitions in the 1 min sit-to-stand (1MSTS) test. Due to restrictions imposed by the COVID-19 pandemic, we were unable to conduct our original study, namely the comparison of the IST with the CPET, the gold standard for exercise capacity, to determine its criterion validity, and with the ISWT to determine its convergent validity. As an alternative, we chose to use the 6MWT, which is a valid and reliable test of exercise capacity for people with chronic lung disease, due to its strong correlations with measures of peak work capacity on a CPET. It is also considered the most widely used field walking test for the assessment of outpatients with COPD. The 1MSTS was also chosen because it is reliable, valid and responsive, and showed a comparable end-exercise cardiorespiratory response to the 6MWT and can induce a similar cardiorespiratory stress to that of CPET.

Reliability refers to the consistency of a measure and its ability to replicate the score from one assessment or rater to another. Measurement error was also considered for reliability, and it is defined as the systematic and random error of a participant’s performance that is not attributed to true changes in the construct to be measured.

A minimum of 50 participants were aimed to be included in this study since this is the sample size suggested by the COSMIN guidelines to determine the construct validity and reliability of measurement instruments with adequate methodological quality.

**Participants**

Patients with COPD were recruited by pulmonologists from two hospitals and two clinics in Portugal. Pulmonologists identified potential participants and ensured the fulfillment of the eligibility criteria. Patients were considered eligible if they had an established diagnosis of COPD based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria—postbronchodilator forced expiratory volume in 1 s (FEV1)/forced vital capacity ratio <70%, were clinically stable over the past month (ie, no hospital admissions or exacerbations), and the presence of an ECG record at rest with no significant changes. Patients were excluded if they had other lung diseases, presence of a significant cardiovascular (eg, symptomatic ischaemic cardiac disease), neurological...
(eg, neuromuscular dystrophy disease), musculoskeletal disease, signs of cognitive impairment or significant risk of fall.

Participants who agreed to participate were contacted by researchers to schedule the appointments for assessment sessions at their homes to provide more information about the study and collect data.

**Data collection**

Participants recruited by pulmonologists were asked to accept two visits for the assessment sessions, performed by one physiotherapist, at their homes, within 5–7 days apart. Lung function tests (spirometry) were collected from all participants and, according to the GOLD guidelines, the airflow limitation of COPD (GOLD I, II, III, IV) of each participant was classified according to the FEV1(%) values. Of these participants, those who performed the 6MWT, according to the American Thoracic Society/European Respiratory Society guidelines, over the last 6MWT, according to the American Thoracic Society/European Respiratory Society guidelines, over the last year were also identified and the respective report (with the main outcome: total distance—6MWD) was also collected.

During the first home visit, sociodemographic (age, sex) and clinical data (medication, comorbidities, smoking status, long-term oxygen, non-invasive ventilation, number of exacerbations, unscheduled consultations, emergency department admission and hospitalisations in the previous year, vital signs, peripheral oxygen saturation—%SpO₂, fatigue and dyspnoea at rest with the modified Borg scale—mBorg) were collected. Anthropometric data (height, weight and body mass index) were collected using a measuring tape and bioelectrical impedance measure—Tanita BC-545 N (Tanita, Amsterdam, The Netherlands). Then, participants performed the 1MSTS once, after a training attempt, with a 5 min period of rest. The first IST (IST-1) was performed after another resting period of at least 20 min to allow for recovery of participant’s vital signs, fatigue, and dyspnoea to their baseline values. During this resting period, patient-reported outcomes measures were collected, namely the Modified Medical Research Council (mMRC) and the COPD Assessment Test (CAT). During the second home visit, 5–7 days afterwards, the vital signs, %SpO₂, fatigue and dyspnoea (mBorg) at rest were collected, and the second IST (IST-2) was performed. The test conditions were similar to the first visit for IST measurements (eg, environment, instructions, same platform to perform the test) and COPD disease stability in participants was also guaranteed.

**Incremental step test**

IST was designed to provide an incremental profile by using a digital recording with timed metronome step cadence, and with a 20 cm tall platform (Max Aerobic step, Mambo, Tisselt, Belgium). The number of levels, and duration of each level (increment) were based on the characteristics of the ISWT.

The original protocol of the ISWT consists of 12 levels; however, as suggested by the literature, we can add more levels to the protocol (total of 15 levels) to allow its future application for other clinical populations, in order to prevent the ceiling effect. Therefore, IST consists in 15 levels, each of 1 min duration. The timed metronome set the step cadence which starts at 10 steps/min and increases 2 steps/min every 1 min, with a step cadence maximum of 38 steps/min (level 15). The maximum test duration is 15 min. Heart rate (HR) and SpO₂ (%) was monitored and registered during the test with a pulse oximeter (PalmSAT 2500 Series, Nonin Medical, Minnesota, USA). The perceived dyspnoea and leg fatigue during the test were also registered with the mBorg scale. The blood pressure was not assessed due to the difficulty of measuring during the stepping.

The criteria to stop the test were: not able to maintain the required step cadence for 10 s, SpO₂ falls to ≤85%, when requested by the participant, or when symptoms were reported (chest pain, intolerable dyspnoea, leg cramps, diaphoresis and a pale or ashen appearance). The main outcome measure of the IST was the total number of steps performed. Maximal step cadence reached, and duration of the test were also collected.

The instructions to perform the IST and a reporting form are available as online supplemental material.

**6 min walk test**

6MWT was performed according to the American Thoracic Society/European Respiratory Society guidelines. The 6MWT is a valid test in people with COPD (moderate to strong correlation with maximum oxygen uptake and peak work on CPET, r=0.40 to 0.93). The 6MWD was the main outcome. HR, SpO₂ (%), perceived dyspnoea and leg fatigue (mBorg scale) were monitored during the test.

**1 min sit-to-stand**

1MSTS was performed on a normal chair available at the participant’s home. Standardised instructions and encouragement were used according to Vaidya et al. This test is a valid test in people with COPD (moderate to strong correlations with peak cycling work capacity and one-repetition maximum, r=0.36–0.63, p<0.05; and positive and strong correlation with the 6MWT, r=0.57 to 0.72, p<0.05). The main outcome measure of the 1MSTS was the total number of repetitions performed. HR, SpO₂ (%), perceived dyspnoea and leg fatigue (mBorg scale) were monitored during the test.

**Patient-reported outcomes measures**

The mMRC and the CAT were used to assess dyspnoea and the impact of the COPD disease, respectively. The mMRC is a 5-point scale with scores ranging between 0 and 4, where higher scores indicate greater dyspnoea severity. Whereas, the CAT is an 8-item scale developed to
assess the impact of COPD through symptoms in patients’ life (cough, sputum, chest tightness, dyspnoea during stair climbing, limitations on home daily activities, confidence to live home, sleep, and energy). Scores range from 0 to 40 and higher scores indicate greater impact of the disease on the patients’ life. The Portuguese versions of the tests are available through the Directorate-General of Health of Portugal website.33

The application of these two instruments, along with the collected information of the number of exacerbations, non-programmed consultations, emergency admission and hospitalisations in the previous year, allowed the application of the GOLD ABCD assessment tool and classification of participants for the assessment of symptoms and risk of exacerbation, according to the GOLD guidelines.4

**Data analysis**

Data analysis was performed using IBM SPSS Statistics V.27.0 (IBM). The level of significance was set at 0.05. Continuous variables were tested for normality with the Kolmogorov-Smirnov and Shapiro-Wilk tests. Descriptive statistics were used, and data are presented as mean±SD, median (percentile 25–75) or frequencies (percentage).

For the assessment of validity, the construct validity34 35 was analysed through the correlation between the number of steps in the best IST and the 6 min walk distance (6MWD), and the number of repetitions in the 1MSTS, using the Spearman correlation coefficient. According to COSMIN recommendations, a ‘positive’ rating to qualify construct validity is determined if the correlation coefficient is equal to or above 0.5.19 In addition, the strength of correlations was classified according to British Medical Journal guidelines: significant correlation coefficients of 0–0.19 as very weak, 0.2–0.39 as weak, 0.4–0.59 as moderate, 0.6 0.79 as strong and 0.8–1.0 as very strong.36

Reliability was determined by intraclass correlation coefficient (ICC) model 2 (two-way random effects), absolute agreement, with a single rater (ICC,2,1), and with 95% CI.37 According to COSMIN recommendations, a ‘positive’ rating to qualify reliability is determined if the ICC value is above 0.70.19 Measurement error was determined calculating the SE of measurement (SEM) and the minimal detectable change at 95% CI (MDC95).24 The SEM was measured according to the following equation:

$$SEM = SD \times \sqrt{1 - ICC}$$

where SD is the SD of the performances obtained from all participants (IST-1 and IST-2). The %SEM was calculated as:

$$\%SEM = \left( \frac{SEM}{mean} \right) \times 100$$

The MDC95 was calculated as follows:

$$MDC95 = 1.96 \times SEM \times \sqrt{2}$$

The %MDC95 was calculated as:

$$\%MDC95 = \left( \frac{MDC95}{mean} \right) \times 100$$

where ‘mean’ is the mean of the performances obtained in IST-1 and IST-2. A %MDC95 of less than 30% was considered acceptable.38

The learning effect was explored using Wilcoxon signed-rank test or paired t-test to compare the performance (number of steps) between the two attempts of the IST. The same tests were used to compare other variable performances (duration, step cadence reached) and physiological response (HR, %SpO2, dyspnoea and

Figure 1 Flow diagram of participants through the study. AECOPD, acute exacerbation of chronic obstructive pulmonary disease; 1MSTS, 1 min sit-to-stand; 6MWT, 6 min walk test.
Sixty participants with COPD were screened to be included in the study. Ten participants were excluded due to: dropped-out for no reason given (n=3), acute infection or post-COVID-19 status (n=2), acute exacerbation of COPD (n=1), presence of a significant musculoskeletal disease (n=2), neurological disease (n=1) and oncological disease (n=1). Therefore, fifty participants were included to assess reliability and validity (with the 1MSTS) of the IST. To assess validity with the 6MWT, only 21 participants (42%) were included since 29 participants (58%) did not perform the 6MWT (figure 1). The main reason for not including these participants was that they did not perform the 6MWT in the previous month and/or performed the test for more than a month.

The characteristics of the fifty participants included in the study are presented in table 1. Most of these participants were males (28 males, 56%), aged 70.8±7.5 years, had moderate airflow limitation (GOLD II, 34 participants, 68%) and belonged to GOLD B group (30 participants, 60%). Twelve participants used long-term oxygen therapy (24%) and fourteen participants used non-invasive ventilation (28%) (table 1).

Table 1  Baseline characteristics of participants

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Eligible participants (n=50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>70.8±7.5</td>
</tr>
<tr>
<td>Sex, male (%)</td>
<td>28 (56)</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>27.1±4.6</td>
</tr>
<tr>
<td>mMRC (total score)</td>
<td>2 (1, 3)</td>
</tr>
<tr>
<td>CAT (total score)</td>
<td>12.8±6.8</td>
</tr>
<tr>
<td>GOLD stages (I, II, II, IV), n (%)</td>
<td>5, 34, 7, 4 (10, 68, 14, 8)</td>
</tr>
<tr>
<td>GOLD group (A, B, C, D), n (%)</td>
<td>10, 30, 0, 10 (20, 60, 0, 20)</td>
</tr>
<tr>
<td>FEV1, %predicted</td>
<td>57.6±16.5</td>
</tr>
<tr>
<td>FEV1/FVC%</td>
<td>56.7±12.1</td>
</tr>
<tr>
<td>Long-term oxygen therapy, n (%)</td>
<td>12 (24)</td>
</tr>
<tr>
<td>Non-invasive ventilation, n (%)</td>
<td>14 (28)</td>
</tr>
<tr>
<td>Cardiac disease</td>
<td>8 (16)</td>
</tr>
<tr>
<td>Arrhythmia</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Heart failure</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>24 (48)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>29 (58)</td>
</tr>
<tr>
<td>OSAS</td>
<td>15 (30)</td>
</tr>
<tr>
<td>Medication, n (%)</td>
<td></td>
</tr>
<tr>
<td>SABA</td>
<td>9 (18)</td>
</tr>
<tr>
<td>LABA</td>
<td>7 (14)</td>
</tr>
<tr>
<td>SAMA</td>
<td>8 (16)</td>
</tr>
<tr>
<td>LAMA</td>
<td>3 (6)</td>
</tr>
<tr>
<td>LABA+LAMA</td>
<td>15 (30)</td>
</tr>
<tr>
<td>LABA+ICS</td>
<td>4 (8)</td>
</tr>
<tr>
<td>LABA+LAMA+ICS</td>
<td>24 (48)</td>
</tr>
<tr>
<td>ICS</td>
<td>8 (16)</td>
</tr>
<tr>
<td>Xanthines</td>
<td>16 (32)</td>
</tr>
</tbody>
</table>

Data are expressed as mean±SD, unless otherwise stated. BMI, body mass index; CAT, COPD Assessment Test; FEV1, forced expiratory volume in 1 s; FVC, forced vital capacity; GOLD, Global Initiative for Chronic Obstructive Lung Disease; ICS, inhaled corticosteroids; LABA, long-acting β2-agonists; LAMA, long-acting muscarinic antagonists; mMRC, Modified Medical Research Council; OSAS, obstructive sleep apnea syndrome; SABA, short-acting β2-agonists; SAMA, short-acting muscarinic antagonists.

RESULTS

Sixty participants with COPD were screened to be included in the study. Ten participants were excluded due to: dropped-out for no reason given (n=3), acute infection or post-COVID-19 status (n=2), acute exacerbation of COPD (n=1), presence of a significant musculoskeletal disease (n=2), neurological disease (n=1) and oncological disease (n=1). Therefore, fifty participants were included to assess reliability and validity (with the 1MSTS) of the IST. To assess validity with the 6MWT, only 21 participants (42%) were included since 29 participants (58%) did not perform the 6MWT (figure 1). The main reason for not including these participants was that they did not perform the 6MWT in the previous month and/or performed the test for more than a month.

The characteristics of the fifty participants included in the study are presented in table 1. Most of these participants were males (28 males, 56%), aged 70.8±7.5 years, had moderate airflow limitation (GOLD II, 34 participants, 68%) and belonged to GOLD B group (30 participants, 60%). Twelve participants used long-term oxygen therapy (24%) and fourteen participants used non-invasive ventilation (28%) (table 1).

Construct validity

The correlation between the number of steps of the best IST and the number of repetitions of 1MSTS was significant, positive, and moderate (ρ=0.50, p=0.020) (figure 2A). However, according to COSMIN recommendations, the correlation coefficient was not equal to or higher than 0.5.

There was a significant difference in the number of steps of the best IST and the 6MWD was also significant, positive and moderate (ρ=0.46, p=0.001) (figure 2B). Despite a smaller number of participants analysed in this correlation, the correlation coefficient achieved the COSMIN recommendations (≥0.5).

Reliability and learning effect

According to COSMIN recommendations, IST showed a high ICC2,1 value (0.96; 95% CI 0.92 to 0.98), for the smaller number of participants analysed in this correlation. The correlation coefficient was not equal to or higher than 0.5.

There was a significant difference in the number of steps of the best IST and the number of repetitions of 1MSTS was significant, positive, and moderate (ρ=0.50, p=0.020) (figure 2A). However, according to COSMIN recommendations, the correlation coefficient was not equal to or higher than 0.5.

There was a significant difference in the number of steps of the best IST and the 6MWD was also significant, positive and moderate (ρ=0.46, p=0.001) (figure 2B). Despite a smaller number of participants analysed in this correlation, the correlation coefficient achieved the COSMIN recommendations (≥0.5).

RESULTS

Sixty participants with COPD were screened to be included in the study. Ten participants were excluded due to: dropped-out for no reason given (n=3), acute infection or post-COVID-19 status (n=2), acute exacerbation of COPD (n=1), presence of a significant musculoskeletal disease (n=2), neurological disease (n=1) and oncological disease (n=1). Therefore, fifty participants were included to assess reliability and validity (with the 1MSTS) of the IST. To assess validity with the 6MWT, only 21 participants (42%) were included since 29 participants (58%) did not perform the 6MWT (figure 1). The main reason for not including these participants was that they did not perform the 6MWT in the previous month and/or performed the test for more than a month.

The characteristics of the fifty participants included in the study are presented in table 1. Most of these participants were males (28 males, 56%), aged 70.8±7.5 years, had moderate airflow limitation (GOLD II, 34 participants, 68%) and belonged to GOLD B group (30 participants, 60%). Twelve participants used long-term oxygen therapy (24%) and fourteen participants used non-invasive ventilation (28%) (table 1).

Construct validity

The correlation between the number of steps of the best IST and the number of repetitions of 1MSTS was significant, positive, and moderate (ρ=0.50, p=0.020) (figure 2A). However, according to COSMIN recommendations, the correlation coefficient was not equal to or higher than 0.5.

There was a significant difference in the number of steps of the best IST and the 6MWD was also significant, positive and moderate (ρ=0.46, p=0.001) (figure 2B). Despite a smaller number of participants analysed in this correlation, the correlation coefficient achieved the COSMIN recommendations (≥0.5).

Reliability and learning effect

According to COSMIN recommendations, IST showed a high ICC2,1 value (0.96; 95% CI 0.92 to 0.98), for the number of steps. Concerning measurement error, SEM and MDC95 were 10.1 steps (%SEM=16.6%) and 27.9 steps (%MDC95=45.8%), respectively. The %MDC95 was considered unacceptable.

There was a significant difference in the number of steps performed between IST-1 and IST-2 (p=0.030), where IST-2 presented a higher median (table 2). Consequently, significant differences were observed in duration (p=0.020), and maximal level achieved (p=0.005) between the IST-1 and IST-2. No differences were found in HR, SpO2%, dyspnoea and leg fatigue in pretest and post-test between the IST-1 and IST-2. Significant differences were found in HR, SpO2%, dyspnoea and leg fatigue before and after the completion of each test (IST-1 and IST-2) (table 2).

DISCUSSION

This study demonstrated that this new IST presented significant and positive correlations with the 6MWT and
1MSTS. Despite these important results, we only reached the sample size recommended by COSMIN through the 1MSTS analysis. However, the correlation coefficient between the IST and 1MSTS was not equal to or higher than 0.5 to consider it as a ‘positive’ quality for construct validity. Therefore, the IST cannot yet be considered a valid test, based on this construct validity, to be used in the assessment of people with COPD. According to the reliability results, a high ICC value, a learning effect, and an ‘indeterminate’ measurement error are shown. This study also demonstrated the feasibility of this test at the home environment, since data collection was performed at participants’ homes and no adverse events were reported.

Regarding the construct validity, the IST showed a moderate correlation with the 6MWT. Although the tests present different modes (walking vs stepping) and profiles (incremental vs self-paced) of testing, this strength of correlation was expected since other step tests applied in people with COPD, including CST and MIST (other IST), presented moderate and strong correlations with the 6MWT (correlation values: 0.56–0.83). As in our study, these correlations were mostly analysed with the performance variables of the tests (number of steps and 6MWD), which support the conceptulation of these step tests, in particular our IST, as important options to assess functional exercise performance. This finding is supported by the fact that guidelines qualify the 6MWT as a more targeted outcome for functional exercise performance. However, we were unable to reach the target sample size for the correlation between IST and 6MWT, although the correlation coefficient reached the COSMIN’s recommendations for a ‘positive’ rating. During the COVID-19 pandemic, health services were reduced with an impact on the number of assessments in outpatients with COPD, and, consequently, a low number of patients who performed the 6MWT during data collection were identified. Therefore, further studies with more participants are necessary to confirm these results and to determine the construct validity of the IST. On the other hand, the target sample size for the correlation between IST and 1MSTS was reached with a moderate correlation, but a ‘negative’ rating for the coefficient correlation was identified (lower than 0.5), according to COSMIN recommendations. Despite these results, to the best of our knowledge, this is one of the first studies to analyse the correlation of a step test and determine its construct validity from a sit-to-stand test.

Regarding reliability, the target sample size was reached. Our study found a ‘positive’ rating for reliability based on the high ICC (0.96; 95% CI 0.92 to 0.98) for the IST, which indicates that this test provides consistent results when the test is applied on different occasions. Other studies that determined the reliability of other IST

**Figure 2** Correlations between the incremental step test and the 1MSTS (A) and 6MWT (B). IST, Incremental step test; 1MSTS, 1 min sit-to-stand; 6MWT, 6 min walk test.

**Table 2** Performance and response of the IST-1 and the IST-2

<table>
<thead>
<tr>
<th></th>
<th>IST-1</th>
<th>IST-2</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>n</em>=50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No of steps</td>
<td>49.5 (25;72)</td>
<td>52 (22; 89.25)*</td>
</tr>
<tr>
<td>Duration (s)</td>
<td>234 (134.5;312.75)</td>
<td>245 (124.25;363)*</td>
</tr>
<tr>
<td>Step cadence (steps/min)</td>
<td>14 (12;18)</td>
<td>15 (12;18)*</td>
</tr>
<tr>
<td>Heart rate (bpm)</td>
<td>77.8±11.5</td>
<td>103.7±16.4†</td>
</tr>
<tr>
<td>SpO₂ (%)</td>
<td>95±1.9</td>
<td>90.3±4.1†</td>
</tr>
<tr>
<td>Dyspnoea (rating)</td>
<td>0 (0; 1)</td>
<td>5 (4; 7)†</td>
</tr>
<tr>
<td>Leg fatigue (rating)</td>
<td>0 (0; 0)</td>
<td>5 (4.75; 6.25)†</td>
</tr>
<tr>
<td>Values are expressed as median (percentile 25–75) or mean±SD.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*P&lt;0.05 for comparisons between the IST-1 and IST-2.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>†P&lt;0.05 for comparisons within tests (pretest vs post-test).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IST, Incremental Step test; SpO₂, peripheral oxygen saturation.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
people with COPD have presented results similar to ours, such as the CST (ICC=0.99) and the MIST (ICC=0.99). Nevertheless, different types of reliability were conducted between studies (ie, within-day vs between-day reliability), especially for the CST, thus, caution should be taken when establishing comparisons. Another finding in our results was the significant difference in the number of steps between IST-1 and IST-2, which indicates a learning effect, suggesting that two tests are required in clinical practice and the result of the second test should be recorded. According to the results for measurement error of the IST, the MDC95 value determined suggests that it is necessary to improve above 27.9 steps to assume that a statistical change in participants’ performance was achieved. Although this cut-off is informative, the calculation and the interpretation of the MDC95 value alone cannot rate the quality of the measurement error as a measurement property. To rate it and to consider it ‘positive’, the MDC95 value must be lower than the minimal important change (MIC), which is defined as the smallest change in the outcome of interest that patients perceive as important, either beneficial or harmful, and that would lead the patient to consider a change in management. However, the MIC was not determined in this study, which rates the IST, for now, as ‘indeterminate’ for measurement error. Therefore, future studies should aim to determine the MIC of the IST (eg, based on PR interventions) in patients with COPD, to determine the quality of this measurement property. The MIC will also provide important information for the interpretability of the IST. Even so, the MDC95% of our study was above the 30% acceptable limit, which can induce a ‘negative’ rating. One explanation for this high MDC95% could be the heterogeneity of groups participants included in our sample, according to the ABCD assessment tool (GOLD A, B, C and D). The ABCD classification appears to be important to discriminate patients with worst outcomes, and therefore, participants from ABCD groups present a heterogeneity of symptoms and exercise capacity levels, despite their stable COPD condition during the study.

There are some strengths and limitations of this study that need to be acknowledged. An important strength is that we tried to assess the measurement properties of a field test to assess people with COPD and to be used in any setting, including at the home environment. Moreover, despite the inability of full compliance, the methodology and sample size used were defined following COSMIN guidelines, which provides general principles for study designs on measurement properties. One important limitation is that we only attempted to assess construct validity, and we did not determine the criterion validity of the IST based on the comparison with the gold standard test to assess exercise capacity—CPET. Future studies should address the assessment of criterion validity by correlating the performance and cardiorespiratory variables, especially the VO2peak, between the IST and the CPET. Additionally, it is important to analyse if this IST can have a maximal cardiorespiratory response.

CONCLUSION

Despite the significant, positive and moderate correlations with the 6MWD and 1MSTS, the inability to fully comply with the COSMIN recommendations to determine measurement properties does not yet allow the IST to be considered a valid test to be used in the assessment of people with COPD. On the other hand, according to COSMIN recommendations, the IST is a reliable test based on its high ICC value. However, a learning effect and an ‘indeterminate’ measurement error are shown.

This study also demonstrated the feasibility of the IST at the home environment since no adverse events were reported during data collection. This test can provide an alternative outcome measure in the assessment of people with COPD to be applied in PR programmes in all settings, including home-based programmes, but further studies are important to determine its measurement properties.

Author affiliations
1Department of Physiotherapy and Center for Rehabilitation Research, School of Health of Polytechnic Institute of Porto, Porto, Portugal
2Healthcare Department, Nippon Gases Portugal, Maia, Portugal
3Pulmonology Department, Centro Hospitalar Barreiro, Montijo, Barreiro, Portugal
4Pulmonology Coordination, Clínica CUF Almada, Almada, Portugal
5Fisioterapia, Matosinhos, Portugal
6Microbiology Research Laboratory on Environmental Health, Institute of Environmental Health, Faculty of Medicine, University of Lisbon, Lisbon, Portugal
7Institute for Preventive Medicine and Public Health, Faculty of Medicine, University of Lisbon, Lisbon, Portugal
8Department of Physiotherapy, Santa Maria Health School, Porto, Portugal

Contributors Conceptualisation: RV and AMM. Methodology: RV, LS, AA, CA, PMS and CC. Validation: CC and AMM. Formal analysis: RV. Investigation: RV, LS, AA, CA, PMS, CC. Resources: CC and AMM. Data curation: CC and AMM. Writing—original draft preparation: RV. Writing—review and editing: RV, LS, AA, CA, PMS, CC. Supervision: CC and AMM. Guarantor: RV. All authors read and agreed to the published version of the manuscript.

Funding This work received financial support from Nippon Gases Portugal to obtain the materials to carry out the field tests.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Consent obtained directly from patient(s).

Ethics approval Ethical approval for this study was obtained from the Ethics Committees of the School of Health - Polytechnic Institute of Porto (Ed134, 13 April 2022).
and license their derivative works on different terms, provided the original work is not modified and/or omissions arising from translation and adaptation or otherwise. BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES


19 Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.
43 Pichon R, Coutraud F, Mialon P, et al. Responsiveness and minimally important difference of the 6-minute Stepper test in


