ESCOLA
SUPERIOR
DE SAÚDE
POLITÉCNICO
DO PORTO

M

MESTRADO
Fisioterapia Cardiorrespiratória

Measurement Properties and Physiological Response Types of Step-tests In Different Clinical Populations – A Systematic Review

Inês Sofia Neves Laranjeira

10/**2021**

ESCOLA
SUPERIOR
DE SAÚDE

Measurement Properties And Physiological Response Types Of Step-tests In Different Clinical Populations - A Systematic Review

Autor

Inês Sofia Neves Laranjeira

Orientador

Professor Doutor António Mesquita Montes / Escola Superior de Saúde – Instituto Politécnico do Porto

Mestre Rui Alves Vilarinho / Escola Superior de Saúde – Instituto Politécnico do Porto

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em **Fisioterapia** – Opção de Cardio-respiratória pela Escola Superior de Saúde do Instituto Politécnico do Porto.

Agradecimentos

Agradeço ao meu corpo e mente, que apesar dos sustos conseguem manter-se de pé contra qualquer adversidade.

Agradeço aos meus pais e irmão, que me compreendem e apoiam incondicionalmente.

Agradeço à Lili, por todos os suspiros, desespero e gargalhadas.

Agradeço às minhas colegas de trabalho, por me terem dado as palavras de incentivo e o silêncio que eu precisei, mas também os risos e abraços, por entre períodos enterrada no computador.

Agradeço à Mixórdia de Temáticas.

Agradeço ao Dr. João Pinto, porque eu também consigo ser uma tardigrada.

Agradeço aos meus orientadores, por me terem ajudado a superar este ano de tese e dado o espaço para conseguir superar todos percalços.

Resumo

Introdução: Avaliação da capacidade de exercício é uma medida clínica importante. A medida *gold standard* é o VO_{2max}, medido por prova cardiopulmonar, que nem sempre é possível aplicar, principalmente em contexto de tratamento domiciliário. Os *step test* são um tipo de exercício de campo com correlação significativa com *gold standard*, rápidos, de baixo custo e fáceis de aprender, no entanto são variados, não permitindo perceber qual o melhor teste a ser usado em cada população. **Objetivos**: Identificar os diferentes *step tests* e respetivo nível de evidência das propriedades métricas (validade fiabilidade e poder de resposta), bem como averiguar as respostas máximas fisiológicas e limitações ao exercício, em cada população. **Bases de dados**: As bases de dados PubMed, MEDLINE, Cochrane Central, PEDRo, LILACS, Web of Science foram utilziadas entre o dia 10 e 30 de julho, 2021. **Métodos**: Foi selecionado qualquer tipo de artigo que reportava propriedades métricas dos step test e respostas fisiológicas correspondentes na avaliação da aptição cardiorrespiratória em qualquer população clínica. Um avaliador retirou toda a informação e avaliou o nível de evidência segundo a COSMIN e a qualidade do desenho de estudo QATAS. Posteriormente foi revisto por dois avaliadores. **Resultados:** Quarenta e três estudos foram incluídos. 2MST, 3MST, 6MST, 6MSrT, aCST, CST, IST, MIST, PST, SST, SFRST, 3STx1, STEP, YMCA+1mHRC foram identificados. O nível de evidência mais comum foi "Baixo" e "Muito Baixo". O melhor nível de evidência foi "Moderado" para validade de critério e fiabilidade teste-retest do 6-Minute Stepper Test em doentes respiratórios (r = 0.48 – 0.63; ICC = 0.92 – 0.95, respetivamente) e cálculo do erro no Adapted Chester Step Test em população saudável (SEM = 12,8). A qualidade do desenho de estudo variou entre "Fraco" e "Moderado". Cinco estudos apresentaram variáveis fisiológicas para avaliação de resposta máxima. Um estudo documentou plateâu do VO_{2max}. Frequência cardíaca e escalas de perceção ao esforço foram as variáveis mais avaliadas. Os step tests apresentam uma limitação ventilatória em doentes respiratórios, uma limitação cardíaca em populações cardíacas e metabólicas e tendem a apresentar uma resposta de limitação por sintomatologia em população saudável. **Conclusão:** O nível de evidência das propriedades métricas dos *step test* é "Baixo" ou "Muito Baixo" limitando a sua aplicação na prática clínica, na avaliação da capacidade de exercício. *Step tests* autocompassados tendem a ter respostas submáximas ao exercício. O 6-Minute Stepper Test e 6-Minute Step são os testes mais utilizados e investigados.

Keywords: Capacidade de exercício, validade, fiabilidade, poder de resposta, respostas cardiorrespiratórias

Abstract

Introduction: Exercise capacity assessment is an important clinical measure. VO_{2max} is the gold standard, measured by cardiopulmonary exercise testing. Nonetheless, it can be difficult to apply, especially in home-based settings. Step tests are a field test type with significant correlation with gold standard, being quick, low cost and easy to learn, however there are multiple protocols, difficulting the choice process in each clinical population. **Aims**: To determine the different step tests, corresponding level of evidence of their measurement properties (validity, reliability and responsiveness) and assess the physiological maximal responses and exercise limitations in any clinical group. **Data sources**: The data sources PubMed, MEDLINE, Cochrane Central, PEDRo, LILACS, Web of Science were searched up between 10th and 30th of July, 2021. **Methods**: Studies of any design that reported any measurement properties of the step test and corresponding physiological responses while assessing exercise capacity in any clinical group were selected. One reviewer collected the data and rated the level of quality by COSMIN and study desing quality by QATAS and was posteriorly reviewed by two other reviewers. **Results**: Forty-three studies were included. 2MST, 3MST, 6MST, 6MSrT, aCST, CST, IST, MIST, PST, SST, SFRST, 3STx1, STEP, YMCA+1mHRC were identified. Level of evidence was mostly "Low" and "Very Low". Best level of evidence was "Moderate" for criterion validity and reliability of the 6-Minute Stepper Test in COPD patients (r = 0.48 – 0.63; ICC = 0.92 – 0.95, respectively) and measurement error on the Adapted Chester Step Test on healthy subjects (SEM = 12,8). Study design quality varied between "Weak" and "Moderate". Only five studies presented all the physiological variables for maximal response assessment. Only one study documented VO_{2max} plateau. Heart rate and rating of perceived exertion were the most commonly evaluated physiological variables. Step tests show a ventilatory limitation on respiratory clinical groups, a cardiac limitation on cardiac and metabolic groups and tend to be more symptom limited on healthy groups. **Conclusion:** The level of evidence of the measurement properties of the step test is "low" to "very low" while assessing exercise capacity, limiting their applicability on clinical practice. Self-paced step tests tend to have submaximal responses while assessing exercise capacity. The 6-Minute Stepper Test and 6-Minute Step test are currently the most frequent and investigated tests.

Keywords: Exercise testing, validity, reliability, responsiveness, cardiorespiratory responses

Index

1. Introduction

"Physical therapy provides services to individuals and populations to develop, maintain and restore maximum movement and functional ability throughout the lifespan. This includes providing services in circumstances where movement and function are threatened by ageing, injury, diseases, disorders, conditions or environmental factors. Functional movement is central to what it means to be healthy", as defined by the World Confederation of Physical Therapy (WCPT, 2011). Physiotherapy and rehabilitation occurs in different settings – inpatient (acute care (hospital), ambulatory (hospital, clinics, specialized centres) and outpatient (community and home-based) (Bott et al., 2009; Geoffrey et al., 2015; Rajan, 2017). Physiotherapists use various approaches to manage functional disorders, being exercise training one of the most common (Taylor et al., 2007) and one of their key skills (John Gormley, 2009). Inherent to exercise prescription is the assessment of physical fitness – muscle strength, flexibility, body composition and cardiorespiratory fitness (CFR) (American College of Sports Medicine, 2010; Tveter et al., 2014).

Maximal oxygen uptake (VO_{2max}) is accepted as the gold standard of CRF. Significant variation in VO_{2max} across populations and fitness levels results primarily from differences in cardiac output. The designation of VO_{2max} implies an individual's true physiologic limit has been reached and a plateau in VO_2 may be observed between the final two work rates of a progressive exercise test (most commonly carried out using a treadmill or cycle ergometer (Huggett et al., 2005)). This plateau is rarely observed in individuals with CVD or pulmonary disease (American College of Sports Medicine, 2010; Arena et al., 2007). Therefore, a test can be considered maximal without the plateau if three of the following criteria are present: elevated respiratory exchange ratio (RER) ≥ 1.15 (Edvardsen et al., 2014), the achievement of >90% of the age-adjusted estimate of maximum heart rate (HR_{max}) (Edvardsen et al., 2014), high postexercise blood lactate levels (≥8 mmol/NL), or the subject's rating of perceived exertion >17 on Borg Scare Rating (6-20) or ≥8 on Modified Borg Scale Rate (0-10) (Edvardsen et al., 2014a; Howley et al., 1995) or Visual Analog Scale (VAS) between 8-10 centimetres (American College of Sports Medicine, 2010; Laveneziana et al., 2021). Consequently, it is understandable that VO_{2max} has a substantial clinical utility for measuring and understanding dysfunction in aging and as a morbidity and mortality predictor (Bruce, 1984; Strasser & Burtscher, 2018; Winett & Carpinelli, 2000)). Despite being gold-standard, it has limitations: requires participants to exercise to the point of volitional fatigue (American College of Sports Medicine, 2010); is limited by musculoskeletal or other impairments; requires higher levels of motivation by the patient; require additional monitoring/emergency equipment and trained medical personnel; and are labour intensive (American College of Sports Medicine, 2010; Sheffield et al., 1969).

In clinical practice, cardiopulmonary exercise testing is not always available, since time is limited, laboratory equipment is unavailable, it may be considered unsafe to exercise at high intensities without medical supervision (Balderrama et al., 2010; Bennett et al., 2016; Sartor et al., 2013). So, it's needed measurement tools with acceptable measurement properties (clinically feasible, readily available and easy to perform in any setting), requiring none or only portable equipment and be time (Tveter et al., 2014), environmental and cost efficient, such as field tests. They are used to evaluate physical and aerobic performance (Tveter et al., 2014) and are considered valid in several populations, specially the ISTW (Monteiro et al., 2014; Wise & Brown, 2005) and 6MWT (Wise & Brown, 2005). These are good predictors of VO_{2max} (Burr et al., 2011; Monteiro et al., 2014; Wise & Brown, 2005; Wyndham, 1967), leading to their frequent use, since being a good measure to assess CRF in healthy and patient groups, especially in cardiorespiratory related pathologies (Prichard et al., 2014; Starobin et al., 2006; Vanhelst et al., 2013). They are scientifically cohesive and are easily reproduced in the community or clinic since little equipment is needed (American College of Sports Medicine, 2010; Sheffield et al., 1969).

Regardless of the applicability of the ISTW and 6MWT for aerobic capacity assessment, home-based practice may not have the needed space to conduct the test, since a mandatory corridor length (10 and 30 meters, respectively) has to be guaranteed. Like so, home-based rehabilitation meets a growing need for physiotherapy treatment, not available in community or clinic-based environments (Stephenson & Wiles, 2000). Specially during COVID-19 pandemic, decentralizing services became a strategy of management and safety, making home-based treatments more necessary (Falvey et al., 2020). With the advantages of proximity and a more interpersonal interaction, environmental disadvantages are also important to have in mind when planning the patient assessment and treatment. Stephenson et al. (2000) evaluated the advantages and disadvantages of the home setting for therapy, having both therapists and patients' perspectives. Both groups mentioned as a disadvantage the limited space and appropriate equipment or equipment needs, making home based treatment potentially less effective (Stephenson & Wiles, 2000). Therefore, what is the best cost-space-effective field test to evaluate aerobic capacity of healthy or patient groups?

Step tests are an inexpensive modality for predicting CFR by measuring HR response to stepping at a fixed rate and/or a fixed step height or by measuring postexercise recovery HR. These require little (a step or a stepper) or no equipment, are easily transportable, stepping skills requires little practice, the test usually is of short duration, and stepping is advantageous for testing bigger sample sizes (American College of Sports Medicine, 2010; Bennett et al., 2016).

The protocols are various and have been studied for their validity in specific populations, more commonly in cardiorespiratory pathologies and healthy populations, both paediatric and adult, over the years (Hamilton & Haennel, 2000; Julio et al., 2017; Mays et al., 2010; Rikli & Jones, 1998; Salbach et al., 2015; Vilarinho, Caneiras, et al., 2021). More recently, a systematic review assessed the current level of evidence of the measurement properties of the step tests in Chronic Obstructive Pulmonary Disease (COPD) (Vilarinho, Caneiras, et al., 2021). Nonetheless there is still lacking evidence that assesses these variables and compares them with several clinical groups, evaluating which steps tests are more frequently used, their clinical applicability, but also their physiological limitations when compared to gold-standard testing.

Therefore, our aim is, primarily, to identify all of the step tests, to critically appraise, compare and summarize the quality of their measurement properties in any clinical group, but also to identify the different physiological responses (maximal or submaximal) according to the variables mentioned above and the consequent limitations of the step tests.

2. Methods

This systematic review was conducted acoording to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Page et al., 2021). It was also registered in the International Prospective Register of Systematic Reviews (PROSPERO), under the code CRD42021260710.

The strategy was conducted according to PICO, described on table 1. The literature search was conducted between the 10th and 30th of July, in the following data bases: PubMed and MEDLINE, Cochrane Central Register of Controlled Trials, Physiotherapy Evidence-Based Database (PEDRo), *Literatura Latino-Americana e do Caribe em Ciências da Saúde* (LILACS), Web of Science, and on the references of other articles assessed throughout the process. The key terms were sellected according to MeSH (Medical Subject Headings) Terms, adapted to each database (since every data base has its own code system) and were the following: "step test", "stepper test", "exercise testing", "aerobic capacity", "cardiopulmonary testing". The key terms were combined in the different data bases with the boolean operators "OR" and/or "AND" (apendix 1). Titles and consequent articles in English or Portuguese, with no age limit, no year intervals, no specific context (community, residents of institutions, or hospital inpatients) and no specific subject of studies (healthy or any clinical population (e.g. respiratory, neuromuscular, cardiac, metabolic, orthopaedic)) were selected. Studies of any design were included, non-published papers or based or posters/abstracts of conference proceedings were excluded. The strategy involved a first screening of all the titles and respective abstracts, followed by the integral reading and analysis (detailed on flow-chart 1).

Table 1: Elegibility Criteria According to PICO Strategy.

	Inclusion	Exclusion
Population	Any population – healthy and clinical population;	NA
Intervention	Step or stepper tests;	Step test without the use of a platform (step or stepper) to step on.
Comparison	CPET or other field tests	NA
Outcome	For the qualitative outcomes will be the different types of step test;	No statistical indicators of measurement properties;

	For the measurement properties we will consider the comparison of the test used with (the criterion and construct validity): 1. cardiopulmonary exercise testing (maximal oxygen uptake), as it is the current gold standard for exercise testing assessment; 2. distance or number of laps obtained in other submaximal tests (e.g. 6MWT, ISWT, respectively). For the physiological responses, it will be considered those mentioned above (variables verified for a test to be considered maximal).	Not mentioning physiological responses.

NA – Non applicable; 6MWT – 6-Minute Walk Test; ISWT – Incremental Shuttle Walking Test.

A step test was defined as a test modality in which the participants are instructed to step up and down on a platform (being a bench, a step, or a stepper device) with a specific yet variant height, in a self or externally (by metronome or audio signal) paced rythim (fixed or incremental rate work intensity). Studies of any design were included, since they reported statistical results for measurment properties (validity, reliability and responsiveness) for assessing exercise capacity in any clinical group. We also considered studies that identified step tests providing maximal responses of physical exertion according to the aims of each test, as showed on table 1. One reviewer (I.S.L.) was responsible for the initial screening and selection of titles and studies, posteriorly. The selected data was reviewed later by a second and third reviewer (R.A.V.; A.M.M).

Data extraction focused on measurement properties, defined according to Consensus-Based Standards for the Selection of Health Measurements Instruments (Prinsen et al., 2018). Validity was defined as the degree to which an instrument measures the construct(s) it purports to measure: criterion validity (the degree to which the scores of an instrument are an adequate reflection of a 'gold standard'; it was considered comparison with cardiopulmonary exercise testing) and construct validity (the degree to which the scores of an instrument are consistent with hypotheses based on the assumption that the instrument validly measures the construct to be measured; it was considered comparisons with six-minute walking test, incremental shuttle walk test, endurance shuttle walk test and other tests available to assess muscle strength) were considered. Reliability was defined as the proportion of the total variance in the measurements which is due to 'true' differences between patients, considering for this review the test-retest reliability (ability of an instrument to produce consistent results when it is used multiple times under nearly equivalent conditions) and measurement error (the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured). Responsiveness was defined as the ability of an instrument to detect change over time in the construct to be measured. In this study it will be considered studies that used the step tests to analyze the effects of any intervention, with changes over time and through comparisons with changes in other outcomes (Prinsen et al., 2018).

It was also assessed for this review the physiological responses of each step test. To assess maximal physiological responses we considered the presence of a plateau of the VO_{2max} or, in case of absence of the plateau, the verification of, at least, three of the following: elevated respiratory exchange ratio (RER) ≥ 1.15 (Edvardsen et al., 2014), the achievement of >90% of the age-adjusted estimate of maximum heart rate (HR_{max}) (Edvardsen et al., 2014), high postexercise blood lactate levels (≥8 mmol/NL), or the subject's rating of perceived exertion >17 on Borg Scare Rating (6-20), ≥8 on Modified Borg Scale Rate (mBorg) (0-10) (Edvardsen et al., 2014; Howley et al., 1995) or Visual Analog Scale (VAS) between 8-10 centimetres (American College of Sports Medicine, 2010; Laveneziana et al., 2021). It was also evaluated the exercise limitation of each step test. A test was considered to have a cardiovascular limitation when heart rate reserve <70% (Bangalore et al., 2006), RER≥1.15, >17 on Borg/≥8 on mBorg/8-10cm on VAS; a ventilatory limitation when breathing reserve < 15%, expiratory flow limitation, signs of hyperinflation (increase in End Expiratory Lung Volume) – when these data is available, HR_{max}< %HR_{max}, >17 on Borg/≥8 on mBorg/8-10cm on VAS; to have a peripherical/other limitation when having signs of leg/back pain, when <16 on Borg/<8 on mBorg/<8 cm on VAS, RER<1.15, or cardiovascular concern (Stickland et al., 2012).

The measurement properties were evaluated using the Consensus-Based Standards for the Selection of Health Measurements Instruments (COSMIN) Risk-of-Bias Checklist (Mokkink, de Vet., et al., 2018), based on a rating score system of four points, where it is provided a classification of the quality of the studies for each question as "very good", "adequate", "doubtful" or "inadequate". For each study, the score per box is obtained by taking the lowest rating of any item in a box. The overall score for each measurement property is determined by the lowest score given for that measurement property. This tool also includes another rating system to evaluate the quality of each measurement property. This rating system consists of giving each criteria: "+" (sufficient), "?" (indeterminate) or "-" (insufficient). This depends on the design used on the studies, but also its methods and outcomes.

After rating the methodological and measurement properties, the modified Grading Recommendations Assessment, Development, and Evaluation (GRADE) approach was applied to grade the level of evidence (based on four factors: risk of bias, inconsistency, imprecision, and indirectness), as "high", "moderate", "low" or "very low" (Mokkink, Prinsen, et al., 2018).

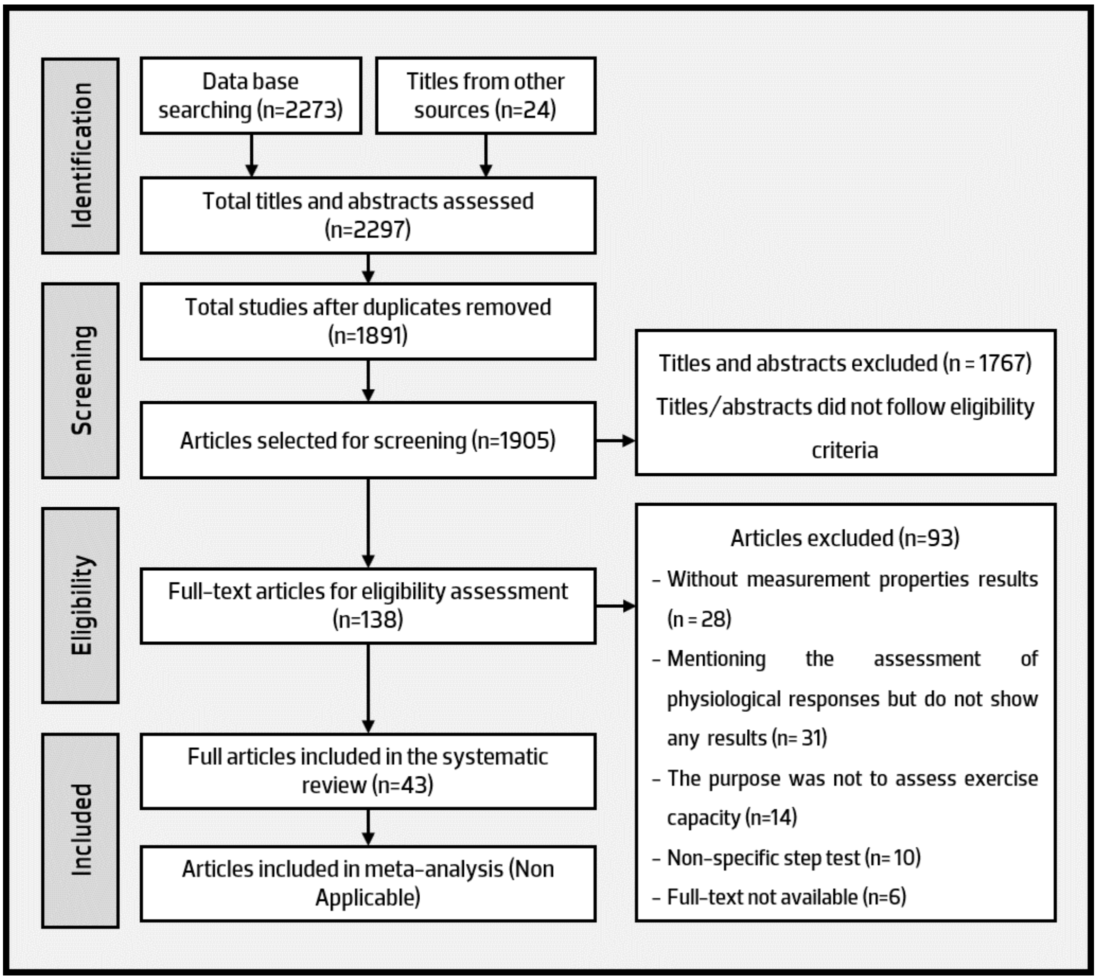
The Quality Assessment Tool for Quantitative Studies (Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies) was used to assess study design quality while assessing physiological responses. This instrument consists of six sections: selection, study design, confounding factors, blinding, data collection method (reliability and validity), and withdrawal, rated as 'strong', 'moderate', or 'weak', based on standardized criteria. These ratings are then combined to attain an overall rating for each study ('strong', no weak ratings; 'moderate', one weak rating; and 'weak', two or more weak ratings) (Armijo-Olivo et al., 2012).

After the review process, the several types of step tests and its general characteristics were sorted. Furthermore, the author names, country and publication year, basic study design (number of participants, mean age, gender distribuition), measurement properties evaluated and physiological responses reported were also selected and organised on different tables, as discriptive synthesis.

3. Results

From the literature search conducted, it was obtained 2283 titles and abstracts. After their screening and duplicate removal, there were 124 full-text articles to assess for eligibility. This assessment resulted in a total a 29 articles to include in this review.

Figure 1: Flow-chart – Literature Search and Analysis Strategy



The detailed information is presented on Figure. 28 studies were written in English and 1 in Portuguese (Ritt et al., 2021). The studies were conducted in Asia – South Korea (n=1) –; Europe – Belgium (n=2), France (n=8), England (n=4), Germany (n=2), Irelant (n=1), Portugal (n=1), Spain (n=1), The Netherlands (n=1)–; North America – Canada (n=3) –; South America – Brazil (n=18), Chile (n=1) –. There were included 1 pilot study, 8 observational studies, 1 cross-over observational study, 1 randomized cross-over study, 16 cross-sectional observational studies, 1 randomized cross-sectional oversational study, 1 prospective randomized study, 4 retrospective studies, 4 prospective studies, 1 retrospective cohort study, 2 prospective cohort studies. Publication date ranges between 1985 and 2021.

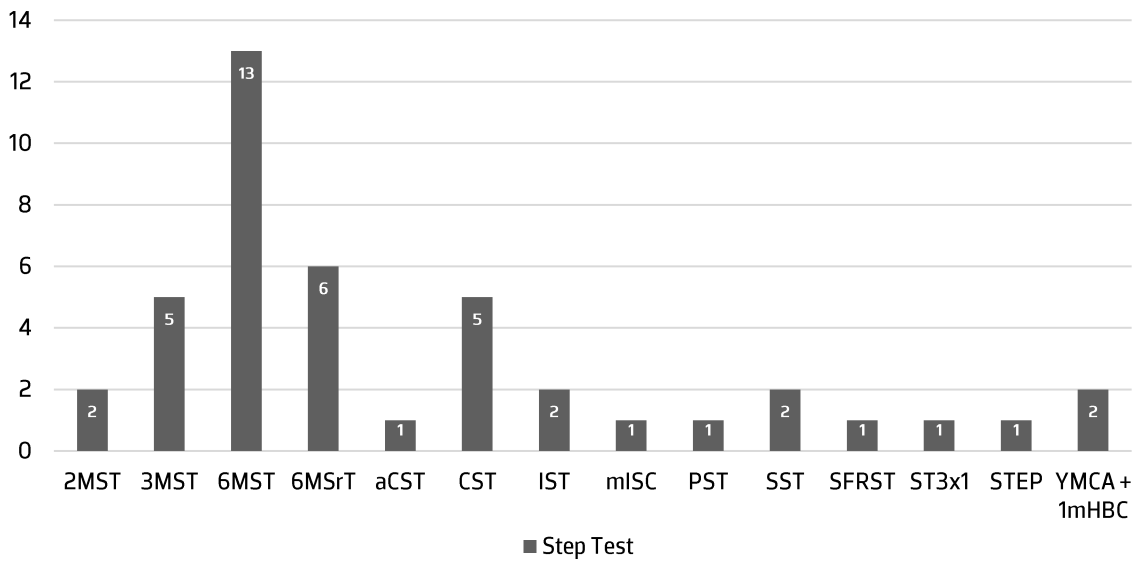
In terms of age limits, only two studies were applied in pediatric populations, both in healthy children (Iturian Barrón et al., 2021; Reychler et al., 2018), leaving the remaining 41 articles executed in adult and geriatric contexts. The total number of individuals participating in all articles selected was 2117 (1222 males and 895 females), with a mean of 60.96 years (youngest population with 8,6±1,8 years and oldest with 69,9±9,0 years). Sample sizes varied in all articles, being the smallest size n=17 and the biggest n=188. 3 studies used only female participants ((Carvalho et al., 2015; Di Thommazo-Luporini et al., 2015; Vilarinho, Mendes, et al., 2021)), 3 studies used only male participants (Simone Dal Corso et al., 2013; Lemanska et al., 2019; Marrara et al., 2012) and 3 studies did not distinguished between genres (Coquart et al., 2016; da Costa et al., 2014; Magalhães et al., 2020).

General study desing information, age groups, genre distribuition, outcomes and other informations are available on appendix 2-3.

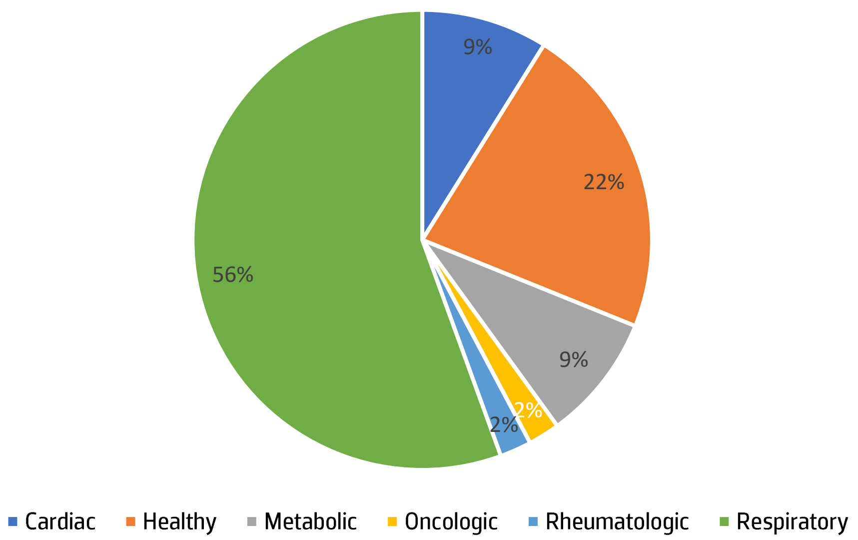
The tests and protocol identified were the following (graphic 1): 2-Minute Step Test (Rikli & Jones, 1999), 3-Minute Step Test (Manual., 2000), 6-min Step Test (S. Dal Corso et al., 2007), 6-minute Stepper Test (Borel et al., 2010), Adapted Chester Step Test (Vilarinho, Mendes, et al., 2021), Chester Step Test (Sykes & Roberts, 2004), Incremental Step Test (de Andrade, De Camargo, et al., 2012), Modified Incremental Step Test (de Andrade, De Camargo, et al., 2012), Paced Step Test (McGavin et al., 1976), Siconolfi Step Test (SF et al., 1985), Standardized Fixed-Rate Step Test (RJ et al., 2003), Step Test 3x1 (Cofre-Bolados et al., 2018), Step Test and Exercise Prescription (STEP) Protocol (Stuckey et al., 2012), YMCA-Step Test, YMCA-Step Test +1-Minute Heart Beat Count (Teren et al., 2016). The identified tests and their characteristics are available on table 2.

There were identified six clinical groups (graphic 2): cardiac diseases (n=4) – cardiovascular disease (n=1), cardiovascular risk factors (n=1), coronary artery disease (n=1), coronary heart disease (n=1); healthy (n=10), methabolic syndrome (n=4) – morbid obesity (n=1), obesity (n=3); oncologic diseases (n=1) – prostate cancer; rheumatology diseases (n=1) – rheumatic arthritis; respiratory diseases (n=9) – acute lung disease (n=1), instestitial lung disease (n=1), COPD (n=19), obstructive sleep apnea (n=1), pulmonary hypertension (n=1).

Graphic 1: number of step tests in this review.



Graphic 2 – distribution of the articles according to clinical groups.



Regarding the measurement properties, 34 articles assessed validity: 18 acessed criterion validity, 11 assessed construct validity, 5 assessed both; 14 assessed reliability, and 9 assessed responsiveness. Criterion Validity was assessed for the 2-Minute Step Test, 3-Minute Step Test, 3-Minute Step Testx1 Minute, 6-Minute Step Test, 6-Minute Stepper Test, Chester Step Test, Incremental Step Test, Modified Incremental Step Test, Paced Step Test, Siconolfi Step Test, Standarded Fixed Rate Step Test, YMCA-

Step Test. This protocols were compared with cardiopulmonary exertion testing on cycle ergometer and treadmill.

Construct validity was included in Chester Step Test, 3-Minute Step Test, 6-Minute Step Test, 6-Minute Stepper Test, Modified Incremental Step Test and they were compared with the 6-Minute Walking Test; Adapted Chester Step Test was compared with the Incremental Shuttle Walking Test, and 6-Minute Spepper Test was compared with 1RM.

Reliability was assessed in 3-Minute Step Test, 6-Minute Step Test, 6-Minute Stepper Test, Adapted Chester Step Test, Chester Step Test, Incremental Step Test, Siconolfi Step Test. All studies presented an appropriate time interval between tests (from 24h to 7 days), except in two reliability analysis on the 3-Minute Step Test and 6-Minute Stepper Test that were conducted on the same day (Beaumont et al., 2019; Bonnevie et al., 2017).

Only 9 studies conducted analysis on responsiveness: eight under pulmonary rehabilitation on COPD patients – 2-Minute Step Test, 3- Minute Step Test, 6-Minute Step Test, 6-Minute Stepper Test after the impleamentation of a rehabilitation program; 3-Minute Step Test to the use of bronchodilation on exertional acute relief and 6-Minute Stepper Test to the use of Neuro Muscular Eletrical Stimulation; another one on cardiovascular rehabilitation program, on the responsiveness of the CST to a cardiac rehabilitation program.

The measurement properties and their level of evidence are available on table 3 and the results on level of evidence assessment are available on appendix 4-6.

Out of the 43 articles available, 34 documented the physiological responses on each step test/protocol used. Nonetheless, only five studies ((Vilarinho, Mendes, et al., 2021), (Vieira et al., 2020), (Carvalho et al., 2015), (Ricci et al., 2019), (Di Thommazo-Luporini et al., 2015)) evaluated the three variables mentioned before. Only one article analysed the blood lactate level (Vilarinho, Mendes, et al., 2021). 28 studies evaluated HR, presenting it as percentage of the HR maximal, except on 11 articles (Beaumont et al., 2019; Bonnevie et al., 2017; Hong et al., 2019; Iturian Barrón et al., 2021; Knight et al., 2014; Lemanska et al., 2019; Munari et al., 2020; Murphy et al., 2005; Pichon et al., 2016; Reychler et al., 2018; Teren et al., 2016; C. de F. Travensolo et al., 2020) where it was analysed the HR_{peak} (peak heart rate) mean-values with standard deviation. RPE was described in 23 articles, 14 using Borg (n=4; varying between 13 and 18) or Modified Borg Scale Rating (n=17; varying between 0 and 6) or Visual Analog Scale (n=2, varying between 2 and 6 cm). 7 studies analysed RER and presented mean values at peak workload.

The results on physiological response assessment are described on table 4.

Table 2 – Step tests identified and corresponding characteristics.

Step test	Studies	Step height	Characteristics			Duration	Workload (Pacing)
			Stages	Initial Work Rate	Increments		
2MST (Rikli & Jones, 1999)	(De Blok et al., 2006) (Ricci et al., 2019)	15,0cm	NA	Maximum number of steps	NA	2 min	Self paced
3MST (Manual., 2000)	(Beaumont et al., 2019) (Borel et al., 2016) (Hong et al., 2019) (Iturian Barrón et al., 2021) (Murphy et al., 2005)	15,0cm or 20,0cm or 20,3 or 30,0cm	NA	Maximum number of steps	NA	3 min	Self paced
6MST (S. Dal Corso et al., 2007)	(Arcuri et al., 2016) (C. de F. Travensolo et al., 2020) (Carvalho et al., 2015) (da Costa et al., 2014) (Di Thommazo-Luporini et al., 2015) (Grosbois et al., 2016) (Magalhães et al., 2020) (Marrara et al., 2012) (Munari et al., 2020) (Pessoa et al., 2014) (Reychler et al., 2018) (Ritt et al., 2021) S. Dal Corso et al., 2007)	15,0cm or 20,0cm	NA	Maximum number of steps	NA	6 min	Self paced
6MSrT (Borel et al., 2010)	(Bonnevie et al., 2017) (Bonnevie et al., 2019) (Coquart et al., 2015) (Coquart et al., 2016) (Fabre et al., 2017) (Pichon et al., 2016)	20,0cm	NA	Maximum number of steps	NA	6 min	Self paced
aCST	(Vilarinho, Mendes, et al., 2021)	20,0cm	15 Stages	Initial level at 15 steps/min	Increments of 2 step/min up until 43 step/min	NA	Incremental (metromone)

(Vilarinho, Mendes, et al., 2021)							
CST (Sykes & Roberts, 2004)	(Reed et al., 2020) (de Camargo et al., 2011) (José & Corso, 2016) (J. Buckley et al., 2004) (Karloh et al., 2013)	15,0; 17,0, 20,0; 25,0; 30,0cm	5 stages (2 minutes each)	15 steps/minute	Increments of 5 steps/min (1 minute)	10 min maximum	Incremental (metromone)
IST (de Andrade, De Camargo, et al., 2012)	(Vieira et al., 2020) (Simone Dal Corso et al., 2013)	20,0cm	Until patient's tolerance	10 steps/minute	Increments of 1 step every 30 s until the patient´s tolerance limit	NA	Incremental (audio signal)
mIST (de Andrade, De Camargo, et al., 2012)	(José & Corso, 2016)	20,0cm	Until patient's tolerance	10 steps/minute	Increments of 1 step every 30 s until the patient´s tolerance limit	NA	Incremental (audio signal)
PST (McGavin et al., 1976)	(Swinburn et al., 1985)	25,0cm	NA	NS	NA	NS	Fixed rate (audio signal)
SST (SF et al., 1985)	(Moore et al., 2013) (Lemanska et al., 2019)	20,0cm	3 stages (3 minutes each)	17 steps/minute	Increments to 26 step/minute and 36 step/minute	6 min	Incremental (metromone)
SFRST (RJ et al., 2003)	(Hansen et al., 2011)	33,0/40,0cm	NA	22.5 steps/min	NA	5 min	Fixed rate (metronome)
ST3x1 (Cofre-Bolados et al., 2018)	(Cofre-Bolados et al., 2018)	20,0/25,0cm	3 stages (minute on; 1 minute off followed by 1 minute of HR counting)	20 cycles/minute	Increments to 32 cycles/minute and maximum cycles/minute the patient can perform	6 min	Incremental (metromone)
STEP (Stuckey et al., 2012)	(Knight et al., 2014)	2 consecutive steps with 20,0cm	20 cycles of stepping	Maximum number of steps	NA	NA	Self paced
YMCA + 1mHBC (Teren et al., 2016)	(Teren et al., 2016) (Beutner et al., 2015a)	30,0cm	NA	20 step cycles/minute	NA	3 min ((2) + 1 minute of HR counting)	Fixed rate (metronome)

2MST - 2-Minute Step Test; 3MST - 3-Minute Step Test; 6MST - 6-min Step Test; 6MSrT - 6-minute Stepper Test; aCST - Adapted Chester Step Test; CST - Chester Step Test; IST - Incremental Step Test; mIST - Modified Incremental Step Test; PST – Fixed Rate Paced Step Test; SST - Siconolfi Step Test; SFRST - Standardized Fixed-Rate Step Test; ST3x1 - Step Test 3x1; STEP - Step Test and Exercise Prescription (STEP); YMCA - YMCA-Step Test; +1mHBC - +1-Minute Heart Beat Count; NA – Non applicable; NS – Not specified; HR – heart rate; cm – centimeters.

Table 3 – Step tests and their measurement properties and respective level of evidence.

Test	Criterion Validity		Construct Validity		Reliability		Measurement Error		Responsiveness	
	Result	Level of evidence	Result	Level of evidence	Result	Level of evidence	Result	Level of evidence	Result	Level of evidence
2MST	r = 0,55 – 0,70 (–)[M]	Low[1]							NA[R,]	Very Low[3]
3MST			r = 0,78[R] (+) r = 0,83 – 0,87[H] (+)	Low[1] – Moderate[2]	ICC = 0.944[H] – 0.964[R] (+)	Low[1]	(?)[R]	Very Low[4]	(?)[R]	Very Low[4]
6MST	r = 0.69[C] (+) r = 0.86[M] (–) r = 0.39 – 0.52[R] (+)	Low[1,2]	r = 0.70[C] (+) r = 0.72[H] (–) r = 0.52 – 0.76 (+)	Low[1,2]	ICC = 0.967[C] (–) ICC = 0.90[H] (–) ICC = 0.80 – 0.98[R] (+)	Low[1]	SEM = 4,27 (–)[R] MDC = 11,8 (–)[R]	Very Low[1,3]	r = 0.34[R]	Very Low[2,4]
6MSrT	r = 0.48 – 0.63 (–)[R]	Moderate[1]	r = 0.41 – 0.72[R] (+)	Low-Very Low[1,2]	ICC = 0.92 – 0.95[R] (+)	Moderate[1,2]	SEM = 17.6;23.5[R] (+)	Very Low[2,3]	r = 0.32[R] (–)	Very Low[2,3]
aCST			r = 0.38 – 0.55[H] (–)	Low[1,3]	ICC = 0.48[H] (–)	Very Low[1,3]	SEM = 12,8[H] (+)	Moderate[1,3]		
CST	r = 0.61[R] (+) r = 0.69[C] (+)	Very Low[1]	r = 0.59 – 0.76[R] (+)	Very Low[1,2]	ICC = 0.99[R] (+)	Very Low[1,3]	(?)[C]	Very Low[1,3]		
IST	r = 0.79 – 0.84[R] (+)	Very Low[1]			ICC = 0.99[R] (–)	Low[1,3]	(?)[R]	Very Low[1,3]		
mIST	r = 0.84[R] (–)	Very Low[2]								
PST	r = 0.74[R] (+)	Very Low[1,3]								
SST	r = 0.69[0] r = 0.79[RH] (?)	Very Low[1,2]			ICC = 0.81[0] ICC = 0.97[RH] (+)	Very Low[1,2]	(?)[0] SEE = 10 – 20[RH] (+)	Very Low[1,2]		
SFRST	r = 0.54[H] (+)	Very Low[1,2]								
ST3x1	r = 0.81[C] (+)	Low[3]								
STEP	r = 0.78[H] (+)	Low[1,3]								

YMCA + 1mHBC	r = 0.64 – 0.862[M] (+)	Low[5]								

2MST – 2-Minute Step Test; 3MST – 3-Minute Step Test; 6MST – 6-min Step Test; 6MSrT – 6-minute Stepper Test; aCST – Adapted Chester Step Test; CST – Chester Step Test; IST – Incremental Step Test; mIST – Modified Incremental Step Test; PST – Fixed Rate Paced Step Test; SST – Siconolfi Step Test; SFRST – Standardized Fixed-Rate Step Test; ST3x1 – Step Test 3x1; STEP – Step Test and Exercise Prescription (STEP); YMCA – YMCA-Step Test; +1mHBC – +1-Minute Heart Beat Count; NA – Non applicable; [1] – small sample n ≤ 50; [2] – moderate small sample n = 50-100; [3] – one study of inadequate quality; [4] – multiple studies of inadequate quality; [5] – inconsistency; [C] – cardiac clinical group; [H] – healthy clinical group; [M] – metabolic clinical group; [O] – oncologic clinical group; [R] – respiratory clinical group; [RH] – rheumatologic clinical group.

Table 4 – Physiological responses in each step test modality.

Reference	Step Test	Clinical Group	Physiological Responses						Exercise Limitations
			VO_{2max} (plateau?)	$\%HR_{max}$ >90%	RER≥1.15	Blood lactate ≥8mol/NL	Rating of perceived exertion >17 Borg; ≥8 mBorg; ≥8cm VAS – Dyspnea	Leg fatigue	
(Reed et al., 2020)	CST	Cardiac – Cardiovascular disease	**NS**						**Step test provoked higher physiological responses when compared to other field tests (RER=1,12±0,8); Without enough information to assess exercise limitation.**
(Cofre-Bolados et al., 2018)	3MSTx1	Cardiac – Cardiovascular risk factors	**NS**						
(Ritt et al., 2021)	6MST	Cardiac – Coronary artery disease	**ND**		1,12±0,8				
(C. de F. Travensolo et al., 2020)	6MST	Cardiac – Coronary artery disease	**ND**	**HR_{paak}** – 117.3±16.3 bpm			**mBorg** – 4.6±2.1		
(Arcuri et al., 2016)	6MST	Healthy	**NS**						**Submaximal ($\%HR_{max}$ = 80 – 90% and RER = 1.01–1.04) and maximal response ($\%HR_{max}$ = 90–94%); Test often complete by reaching the end or individual's maximum, without any specific limitation.**
(Beutner et al., 2015b)	YMCA	Healthy	**ND**						
(Hansen et al., 2011)	SFRST	Healthy	**ND**	86 ± 12%			**Borg** = 13 ± 2.		
(Hong et al., 2019)	3MST	Healthy	**ND**	**HR_{peak}** – 116,7±15,0 bpm to 137,5±17,8 bpm					
(Iturian Barrón et al., 2021)	3MST	Healthy	**ND**	**HR_{peak}** – 147±2,3 bpm				**mBorg** – 4,13±2,2	
(J. Buckley et al., 2004)	CST	Healthy	**$\%VO_{2max}$** – 65,6±9,3	81,1±11,2%			**Borg** – 14,2±2,0		
(Knight et al., 2014)	STEP	Healthy	**ND**	**HR_{peak}** – 174±21 bpm					

(Reychler et al., 2018)	6MST	Healthy	**ND**	**HR_{peak}** – 112,9±25 bpm				**VAS** – 5,40 (0,5–10,0mm)	
(Carvalho et al., 2015)	6MST	Healthy (LN) and Metabolic – obese (OB)	**LN** **VO_{2peak} (ml/kg/min)** – 24,9±3,7 **OB** **VO_{2peak} (ml/kg/min)** – 17,6±3,1	LN: 92±8% OB: 90±9%	LN 1,04±0,09 OB – 1,07±0,10		LN **mBorg** = 3 (0,5–8) OB **mBorg** = 3,5 (0–10)	LN **mBorg** = 3 (0–7) OB **mBorg** = 2 (0–10)	
(Karloh et al., 2013)	CST	Healthy and Respiratory – COPD	**ND**	Control group – 83,5±14,9%; COPD group – 70,5±10,5%;			**ΔBorg:** Control group – 2,6±2,1%; COPD group – 2,6±2,7%.		
(Vilarinho, Mendes, et al., 2021)	aCST	Healthy sedentary	**ND**	94.0 ± 6.5%		11.3 ± 4.5	**Borg** = 18.4 ± 1.5.		
(Ricci et al., 2019)	2MST	Metabolic – Obesity and morbid obesity	**VO_2 (mL/kg/min)** – 12,5 (11,8–13,2)	74,3 (70,9–77,7)%	1,01 (0,96–1,06)		**mBorg** – 2 (0–7)	mBorg – 0 (0–6)	Shorter duration tests evoke submaximal responses (RER = 1.01; $\%HR_{max}$ = 74.3%).
(Teren et al., 2016)	YMCA	Metabolic – Overweight	**VO_{2max} (mL/kg/min)** = **26.4 (25.0–28.1)**	**HR_{peak}** – 120 (103–138) bpm			**Borg** – 14 (12–15)		
(Lemanska et al., 2019)	SST	Oncologic – Prostate cancer	**ND**	**HR_{max}** – 141,6±20,7 bpm (~$\%HR_{max} \geq 90\%$)					Step tests evoked maximal response on HR, and near maximum VO_{2max} to CPET. Subjects seem to reach maximum by exhausting cardiac reserve.
(José & Corso, 2016)	**(CST/MIST)**	Respiratory – Acute lung disease	**ND**	76±11%/76±12%			4 (3–7)/4 (3–7)	mBorg = 5 (3–7)/5 (3–7)	Tendency to exhaust breathing reserve (no available information to

(de Camargo et al., 2011)	CST	Respiratory – COPD	**ND**	CST 1 – 80±10; CST 2 – 80±13;			**mBorg:** CST 1 – 4 (1-10); CST 2 – 4 (0–9)	**mBorg:** CST 1 – 4 (2-10); CST 2 – 5 (0–10).	confirm), having submaximal HR, RER and REP, which is verified. 6MST and IST seem to provoke higher $\%HR_{MAX}$ and RER, near maximal. 6MST elicited maximal response, having described a plateau in one study.
(Beaumont et al., 2019)	3MST	Respiratory – COPD	**ND**	HR_{peak} 112,9±13,4 bpm			**(mBorg)** – 5,0±2,2	**(mBorg)** – 3,4±2,4.	
(Bonnevie et al., 2017)	6MSrT	Respiratory – COPD	**ND**	**HR_{peak} = 120,6 ±.15 bpm**					
(Bonnevie et al., 2019)	6MSrT	Respiratory – COPD	**NS**						
(Borel et al., 2016)	3MST	Respiratory – COPD	**ND**	**83–91%**			**mBorg) – 2,3 – 5,9**	**mBorg) – 2,5 – 4,9**	
(Pessoa et al., 2014)	6MST	Respiratory – COPD	**NS**						
(Coquart et al., 2015)	6MSrT	Respiratory – COPD	**ND**	72,4±3,4 70,4±3,3			mBorg = 3,8±0,6; mBorg = 3,8±0,6	mBorg = 4,1±0,7 mBorg = 3,7±0,5	
(Coquart et al., 2016)	6MSrT	Respiratory – COPD	**NS**						
(De Blok et al., 2006)	2MST	Respiratory – COPD	**NS**						
(Fabre et al., 2017)	6MSrT	Respiratory – COPD	**ND**		1,02±0,14				
(Grosbois et al., 2016)	6MST	Respiratory – COPD	**NS**						
(da Costa et al., 2014)	6MST	Respiratory – COPD	**NS**						
(Marrara et al., 2012)	6MST	Respiratory – COPD	**ND**				**mBorg (experimental/ control) – 1/0.5 (0,5–5,0)**		
(Munari et al., 2020)	6MST	Respiratory – COPD	**Described VO_{2max} plateau**	**HR_{peak} – 119 ± 12.3 ($\%HR_{max}$ ≈ 81.5%)**					

(Murphy et al., 2005)	3MST	Respiratory – COPD	**ND**	**HR$_{peak}$ Experimental – pre: 112±13; post: 116±13 Control – pre: 118,8±15; post: 113±10**			**mBorg Experimental – pre: 5,5±2,55; post: 4,9±1,76 Control: pre – 5,1±3,24; post: 5,3±3,6**		
(Pichon et al., 2016)	6MSrT	Respiratory – COPD	**ND**	Pre RP HR$_{peak}$ – 117.9 ± 15.8 bpm Post RP HR$_{peak}$ – 118.9 ± 17 bpm			Pre RP mBorg - 5.4 ± 2.0 Post RP VAS – 4.9 ± 2.1 cm	Pre RP mBorg - 4.9 ± 2.1 Post RP VAS – 4.5 ± 2.0 cm	
(Simone Dal Corso et al., 2013)	IST	Respiratory – COPD	**IST-1** **VO$_{2max}$** – 1,19±0,39L/min **IST-2** **VO$_{2max}$** – 1,20±0,40L/min	IST-1 88±13% IST-2 86±12%			IST-1 mBorg = 5,4±2,3 IST-2 mBorg = 5,1±2,2	IST-1 mBorg = 5,5±2,6 IST-2 mBorg = 5,1±2,7	
(Swinburn et al., 1985)	PST	Respiratory – COPD	**VO$_{2max}$** = **1010±190 (mL/min)**						
(S. Dal Corso et al., 2007)	6MST	Respiratory – Interstitial lung disease	**VO$_{2max}$ (L/min)** = 0.98±0.22	84.4±8.4	0.95±0.06		**mBorg** = 4 (0-10)	**mBorg** = 4.5 (0-10	
(Di Thommazo-Luporini et al., 2015)	6MST	Respiratory – Obesity	**VO$_{2peak}$** – 17,8±3,6 mL/kg/min	87±9	1,03±0,09		**mBorg** – 3 (0;10)	**mBorg** – 3(0;100)	
(Magalhães et al., 2020)	6MST	Respiratory – OSA	**ND**	74,2±11,8				**mBorg** – 5,1±1,9	
(Vieira et al., 2020)	IST	Respiratory – Pulmonary hypertension	**VO$_{2peak}$** – 1018±283 (mL/min)	84±7%	1,11±0,11		**mBorg** – 5±3	**mBorg** – 6±2	

(Moore et al., 2013)	SST	Reumatologic – Reumathoid arthritis	**ND**	**visit 1** – 73±8%; **visit 2** – 75±8%.					Submaximal response in $\%HR_{MAX}$ without any other information to assess exercise limitation.

2MST – 2-Minute Step Test; 3MST – 3-Minute Step Test; 6MST – 6-min Step Test; 6MSrT – 6-minute Stepper Test; aCST – Adapted Chester Step Test; CST – Chester Step Test; IST – Incremental Step Test; mIST – Modified Incremental Step Test; PST – Fixed Rate Paced Step Test; SST – Siconolfi Step Test; SFRST – Standardized Fixed-Rate Step Test; ST3x1 – Step Test 3x1; STEP – Step Test and Exercise Prescription (STEP); YMCA – YMCA-Step Test; +1mHBC – +1-Minute Heart Beat Count; VO_{2max} – maximal oxygen uptake; HR_{max}– maximal heart rate; $\%HR_{max}$ – percentage of maximal heart rate; RER – respiratory exchange ratio; Borg – Borg Exertion Scale; mBorg – Modified Borg Exertion Scale; VAS – Visual Analog Scale; $\%VO_{2max}$ – percentage of VO_{2max}; VO_{2peak} – peak of maximal oxygen uptake; OSA – Obstructive Sleep Apnea; mL – millilitres; kg – kilogram; min – minute; bpm – beats per minute; ND – Not described; NS – Not Specified.

4. Discussion

The aim of this systematic review was to primarly identify the several step tests currently used in clinical practice in all population and report their measurment properties, but also to analyse the type of physiological response and test exercise limitations induced on the subjects according to each clinical population.

In this review we identified 14 different step tests and they differ externally – fixed rate or incremental, self paced or paced by external signal (audio or metronome) – or internally – various step heights for the same protocol, adapting to height or sex –. The lack of standardized equipment use in between the different testing methods demonstrate that step tests are still more used in investigation rather in practice, since there's no specific guidelines. Another factor corroborating this statement is the percentage of each step test type found in this review. Six of the 14 diferrent step tests (Adapted Chester Step Test, Modified Incremental Step Test, Paced Step Test, Standard Fixed Rate Step Test, 3-Minute Step Test with 1 Minute HRC, STEP Protocol) only had one article available on this review, demonstrating lack of applicability on investigation and practice, and consequent difficulty in assessing their measurement properties, also a finding in a similar systematic review that evaluated the measurement properties of step tests in COPD patients (Vilarinho, Caneiras, et al., 2021).

The level of evidence of the measurement properties evaluated on each step test ranged mostly between "Low" and "Very Low", according to the Modified Grading of Recommendations Assessment, Development, and Evaluation approach (GRADE). This results is mostly caused by small samples of participants in the articles, having just one article on the subject, several articles with "inadequate" ou "doubtful" methodological quality (Mokkink, Prinsen, et al., 2018). Nonetheless, it is important to adress that, despite our efforts in rating methodological quality of the articles, these results can't be taken as absolute since we selected measurement properties from articles that were not specifically designed for that porpuse.

Overall, validity was the measurement property more frequently assessed, being reported in all step tests. Criterion validity was most frequent measurement property in the studies analysed, with comparison with gold-standard CPET with cycle ergometer and treadmill protocols being used, except for the Adapted Chester Step Test. Its level of evidence ranged between "Low" and "Very Low", due to the fact of inconsistencies on the studies and tendency for small sample sizes (<50 participants). Despite that, the results, overall, show good correlation between step test outcomes and CPET outcomes, with no difference between step test modality. Construct validity was also commonly used in this review, since it compared the step tests with field tests, widely used in investigation. All studies assessing construct validity used 6-Minute Walking Test, except Vilarinho et al. (2020) which compared the Adapted Chester Test with the Incremental Shuttle Walking Test (ISWT), Karloh et al (2013), which compared Chester Step Test with ISWT. This choice seems appropriate, because even though they involve different activities, they are similar in nature because both progressively increase the load imposed, which is announced by a signal sound, and are limited by the symptoms of the patient (Karloh et al., 2013). Also, Bonnevie et. al (2019) compared 6-Minute Stepper Test with 1 repetition maximum (1RM), showing significant correlation between them (r=0.41, p<0.001), showing that 6MSrT can have a moderate power to extrapolate strength in COPD patients, but not be used as an indirect measure to assess strength training in this population.

Reliability methodological assessment was more heterogenous, being most step test rated as "Low" or "Very Low", except 6-Minute Stepper Test, with a "Moderate" rating, in respiratory population. This test was shown to have a strong correlation between repetitions (ICC = 0.92 – 0.95), being this result explained by the learning effect. Since step tests are easy to understand and perform it is understandable that leaning effect can affect the performance of the consequent repetitions (Coquart et al., 2015). The poor level of evidence on the remaining tests reflects the lack of pattern of the test-retest waiting period. The second testing day varied between tests from 30 minutes to 7 days, being the most commonly used period the 24-48 hours. Short period between retesting could indicate memory effects from the participants, having a greater performance with compared to the first repetition. Thus, the retest period should be long enough that the participants can't remember or be influenced by the first test (American College of Sports Medicine, 2010).

Responsiveness was poorly evaluated overall, since only 9 studies aimed to assess it, small samples sizes were included and lack of methodological quality on the study design. Several studies failed to present statistical values of the measurement error and sample size, or appropriate statistical measures for the property aimed to be evaluated in the first place. Most studies were described to detect differences after a rehabilitation program, being in cardiac rehabilitation or respiratory, but did not present any data corroborating that statement.

An important find is the presence of measurement properties of the 6-Minute Step Test and 6-Minute Stepper Test, having reported values on all the properties we assessed, besides being the most used step tests. Both have strong correlation with 6-Minute Walk Test, seem to be reliable and have moderate response to rehabilitation programs, being cardiac or respiratory. 6-Minute Step Test and Stepper Test are fast, self-paced, don't require much equipment and have similar commands and encouraging statements to 6MWT. 6MST has been validated in several articles to be a good predictor of exercise capacity in correlation with 6MWT, in healthy, respiratory and cardiac groups (de Andrade, Cianci, et al., 2012; C. Travensolo & Polito, 2014). Also 3-Minute Step Test, as the third most common step test in this review, also self-paced, reporting good validity assessing exercise capacity. Chester Step Test, was the mostly used incremental test on these studies.

In quality of the study design, the ratings varied between "Weak" and "Moderate". The lack of "Strong" evidence on this systematic review is related to the lack of randomization, lack of blindness of the evaluators, lack of blindness of the population being studied, sample selection by convenience and heterogeneity in participants. It is also important to mention, the lack of rigor in writing the experimental design, since several studies didn't describe the step test (fundamentals, application, criteria of termination of the test), only giving a general idea of the test concept.

Physiological responses evaluation showed a significant lack of quality in their evaluation. Most studies did not report physiological responses as a primary outcome, being most of them only shown in descriptive values in the results. This fact is explained by the lack of information available for direct assessment of these responses. Despite not being targeted as main outcomes in most evidence, most of the studies evaluated %HR and RPE through Borg, mBorg or VAS. Only one study (Vilarinho, Mendes, et al., 2021) evaluated three of the 4 variables for a test to be considered maximal. Some studies presented values of $VO_{2max/peak}$. Only one study documented VO_2 plateau. One conclusion after analysing the studies is that most experimental designs chose to evaluate HR and RPE, being these two variables easy to monitor and register, but also low cost, when compared to lactate blood concentration analysed in only 1 study. Some studies presented values for RER, but overall, it was not documented, which shows a lack of methodological quality since RER is the second most important factor to assess for maximal responses. The rise in RER during heavy exercise is caused by an imbalance between the production and the elimination of lactic acid, because of the increase in the buffering of lactate. In addition, as CO_2 is generated from muscle work, the rise in ventilation increases the RER (Edvardsen et al., 2014).

Overall data available was not sufficient to assess if step tests can be considered maximal testing. Nonetheless, it showed important information to make some conclusions.

Step tests tend to have an overestimation of VO_{2max} when is directly measured. This fact is explained by the bigger recruitment of muscle fibres and contraction regiment during stepping when compared to walking (Chéhère et al., 2016; Vilarinho, Mendes, et al., 2021).

4.1. Step-test in Healthy Groups

The following test were applied: 3-Minute Step Test, 6-Minute Step Test, Adapted Chester Step Test, Chester Step Test, Standardized Fixed Rate Step Test, STEP Protocol, YMCA-Step Test. In general, these tests presented moderate to strong correlation to distance in 6MST and VO_2 in CPET. Only Adapted Chester Test induced maximal responses. In healthy subjects, step tests seem to induce bigger responses in HR, but not in RPE Borg. The results on Adapted Chester Step Test could be on physiological responses could be explained by the incremental nature of the test, inducing more stress over time, without the risk of plateau.

4.2. Step-test in Respiratory Groups

It includes the tests: 3-Minute Step Test, 6-Minute Step Test, 6-Minute Stepper Test, Chester Step Test, Incremental Step Test, Modified Incremental Step Test. These tests presented a moderate to strong correlations being the Chester Step Test and Incremental Step Test the best test correlating with gold-standards. All the tests that analyzed reliability presented good results. In respiratory clinical groups, none of the step tests induced maximal responses. The most affected physiological variable was $\%HR_{max}$, RER and Borg for leg fatigue. It is clear that respiratory patients, having the ventilatory capacity affected have a decrease in muscular metabolism by the imbalanced oxygen transportation. When large group muscles are elected to work, since these patients have already a functional deficit, leg fatigue is one of the most affected variables in step test, more than dyspnea. Therefore, step tests in respiratory patients seem to affect more the functional profile of the patient and not the ventilatory deficiency.

4.3. Step-test in Cardiac Groups

There were identified 4 step tests in this clinical group: 6-Minute Step Test, Chester Step Test, Step Test 3x1. These tests presented a good correlation to VO_{2peak} measured by CPET. 6-Minute Step Test had a good correlation to the 6MWT distance and was reliable. Nonetheless, this clinical group was the poorest in physiological responses, having 2 studies failing to mention them. 6-Minute Step Test induced an increase in HR_{peak} and Borg for dyspnea, and a near maximal RER.

4.4. Step-test in Metabolic Groups

Obesity was considered a metabolic disorder (Eckel et al., 2010). Thus, the following step steps were identified: 2-Minute Step Test, 6-Minute Step Test, YMCA-Step Test. 2-Minute Step Test and 6-Minute Step test presented good correlation with VO_2 CPET, being valid to assess maximal workload and oxygen intake. None of the studies induced maximal physiological responses. Analyzing the responses obtained it is clear that in obese patients step tests tend to exert more cardiac responses, by increasing HR. In 6-Minute Step Test, one study demonstrated almost maximal % HR_{max} by achieving almost 87%., which in some studies is considered a maximal response.

4.5. Step-test in Rheumatologic and Oncologic Groups

Siconolfi Step Test was identified for these two clinical groups. This test showed good correlation to VO_{2peak}. Also good reliability. In terms of physiological responses, only HR was assessed for both clinical groups, having a near maximal response in oncologic clinical group, but under maximal values for HR_{max} with around 75%. Alone these values don't allow for extrapolations and conclusions if the step tests can be considered maximal or induce maximal responses in these populations.

Step tests and their considerations for each clinical group is available on table 8.

5. Study limitations and future sugestions

To our knowledge, this is the first systematic review to identify and assess measurement properties and physiological responses in the different step tests in any clinical group. It's also valuable for the inclusion of new protocols and new stepping equipment such as the stepper and double step. We also included studies that evaluated measurement properties without that being the primary outcome of the study making the step tests identified list broader and more diverse, adding value to this review.

It also has limitations. First of all, the study screening, assessment and evaluation was made by one evaluator, increasing the chance of evaluator bias. It was only included studies in Portuguese and English and studies that evaluated measurement properties but did not document physiological responses.

It is important to continue the research on step tests, specially assessing their behavior and response in home-care settings, since it is when this step tests might be more appropriate, since they don't need much equipment and space. It is also important to continue research in this subject with uniform samples, in size and characteristics, making them more reliable. Also important for the future to assess their responsiveness in different settings.

6. Conclusion

According our findings, step tests can differ in stepping frequency, bench height, test duration and number of test stages. They can be fixed, incremental or self-paced, with a step or a stepper. Incremental step-tests have either the step box height or stepping rate increased in an incremental manner; these tests also tend to take longer to complete and elicit more physical stress. Fixed (or single stage) step tests take shorter duration and use heart rate recovery to estimate VO_{2max} and they may produce a less accurate estimation of cardiorespiratory fitness.

It is important to mention that, incremental/fixed rate/self paced step tests induce different results and measures. Incremental tend to be symtom limited, while self paced tend to be more functional having in mind activities of daily life. Their use should be taken into thought while deciding which step test to use. If the target population is respiratory, is natural that self paced tests could be more easily used to assess exercise capacity, since patients have a natural ventilatory limitation, while which healthy subjects, a more sympton limited test, such has Chester Step Test, can be used to assess more variables, besides exercise capacity.

This systematic review identified 14 different step tests: 2MST; 3MST; 6MST; 6MSrT; aCST; CST; IST; mIST; SST; SFRST; ST3x1 ; STEP; YMCA +1mHBC.

Their general level of methodological quality ranged between "Very Low" and" Moderate", being the validity the measurment property with better quality. The level of quality of the study designs ranged between "Weak" and "Moderate", having the worst quality regarding "Blinding".

There was identified six different clinical groups: Healthy, Respiratory, Cardiac, Metabolic, Rheumatological and Oncologic. All have step tests with good correlation with gold-standard or a measured considered gold-standard. Yet, none of the step tests induced maximal physiological responses in the participants, except Adapted Chester Step Test on young healthy adults.

References

American College of Sports Medicine. (2010). ACSM's Guidelines for Exercise Testing and Prescription. In *Lippincott Williams & Wilkins*.

Arcuri, J. F., Borghi-Silva, A., Labadessa, I. G., Sentanin, A. C., Candolo, C., & Di Lorenzo, V. A. P. (2016). Validity and reliability of the 6-minute step test in healthy individuals: A Cross-sectional study. *Clinical Journal of Sport Medicine*, *26*(1), 69–75. https://doi.org/10.1097/JSM.0000000000000190

Arena, R., Myers, J., Williams, M. A., Gulati, M., Kligfield, P., Balady, Collins, E., & Fletcher, G. (2007). Assessment of functional capacity in clinical and research settings: a scientific statement from the American Heart Association Committee on Exercise, Rehabilitation, and Prevention of the Council on Clinical Cardiology and the Council on Cardiovascular N. *Circulation*, *116*(3), 329–343. https://doi.org/10.1161/CIRCULATIONAHA.106.184461

Armijo-Olivo, S., Stiles, C., Hagen, N., Biondo, P., & Cummings, G. (2012). Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *Journal of Evaluation in Clinical Practice*, *18*(1), 12–18. https://doi.org/10.1111/J.1365-2753.2010.01516.X

Balderrama, C., Ibarra, G., De La Riva, J., & López, S. (2010). Evaluation of three methodologies to estimate the VO2max in people of different ages. *Applied Ergonomics*, *42*(1), 162–168. https://doi.org/10.1016/j.apergo.2010.06.017

Bangalore, S., Yao, S. S., & Chaudhry, F. A. (2006). Comparison of Heart Rate Reserve Versus 85% of Age-Predicted Maximum Heart Rate as a Measure of Chronotropic Response in Patients Undergoing Dobutamine Stress Echocardiography. *The American Journal of Cardiology*, *97*(5), 742–747. https://doi.org/10.1016/J.AMJCARD.2005.09.111

Beaumont, M., Losq, A., Péran, L., Berriet, A. C., Couturaud, F., Le Ber, C., & Reychler, G. (2019). Comparison of 3-minute Step Test (3MStepT) and 6-minute Walk Test (6MWT) in Patients with COPD. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, *16*(3–4), 266–271. https://doi.org/10.1080/15412555.2019.1656713

Bennett, H., Parfitt, G., Davison, K., & Eston, R. (2016). Validity of Submaximal Step Tests to Estimate Maximal Oxygen Uptake in Healthy Adults. *Sports Medicine*, *46*(5), 737–750. https://doi.org/10.1007/s40279-015-0445-1

Beutner, F., Ubrich, R., Zachariae, S., Engel, C., Sandri, M., Teren, A., & Gielen, S. (2015a). Validation of a brief step-test protocol for estimation of peak oxygen uptake. In *European Journal of Preventive Cardiology* (Vol. 22, Issue 4, pp. 503–512). https://doi.org/10.1177/2047487314533216

Beutner, F., Ubrich, R., Zachariae, S., Engel, C., Sandri, M., Teren, A., & Gielen, S. (2015b). Validation of a brief step-test protocol for estimation of peak oxygen uptake. *European Journal of Preventive Cardiology*, *22*(4), 503–512. https://doi.org/10.1177/2047487314533216

Bonnevie, T., Allingham, M., Prieur, G., Combret, Y., Debeaumont, D., Patout, M., Cuvelier, A., Viacroze, C., Muir, J.-F., Medrinal, C., & Gravier, F.-E. (2019). The six-minute stepper test is related to muscle strength but cannot substitute for the one repetition maximum to prescribe strength training in patients with COPD. *International Journal of Chronic Obstructive Pulmonary Disease*, *14*, 767. https://doi.org/10.2147/COPD.S193585

Bonnevie, T., Gravier, F. E., Leboullenger, M., Médrinal, C., Viacroze, C., Cuvelier, A., Muir, J. F., Tardif, C., & Debeaumont, D. (2017). Six-minute Stepper Test to Set Pulmonary Rehabilitation Intensity in Patients with COPD–A Retrospective Study. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, *14*(3), 293–297. https://doi.org/10.1080/15412555.2017.1303040

Borel, B., Fabre, C., Saison, S., F, B., & Grosbois, J.-M. (2010). An original field evaluation test for chronic obstructive pulmonary disease population: the six-minute stepper test. *Clinical Rehabilitation*, *24*(1), 82–93. https://doi.org/10.1177/0269215509343848

Borel, B., Wilkinson-Maitland, C. A., Hamilton, A., Bourbeau, J., Perrault, H., Jensen, D., & Maltais, F. (2016). Three-minute constant rate step test for detecting exertional dyspnea relief after bronchodilation in COPD. *International Journal of Chronic Obstructive Pulmonary Disease*, *11*(1), 2991. https://doi.org/10.2147/COPD.S113113

Bott, J., Blumenthal, S., Buxton, M., Ellum, S., Falconer, C., Garrod, R., Harvey, A., Hughes, T., Lincoln, M., Mikelsons, C., Potter, C., Pryor, J., Rimington, L., Sinfield, F., Thompson, C., Vaughn, P., & White, J. (2009). Guidelines for the physiotherapy management of the adult, medical, spontaneously breathing patient. *Thorax*, *64*(SUPPL. 1). https://doi.org/10.1136/thx.2008.110726

Braghieri, H. A., Kanegusuku, H., Corso, S. D., Cucato, G. G., Monteiro, F., Wolosker, N., Correia, M. de A., & Ritti-Dias, R. M. (2021). Validity and reliability of 2-min step test in patients with symptomatic peripheral artery disease. *Journal of Vascular Nursing*, *39*(2), 33–38. https://doi.org/10.1016/j.jvn.2021.02.004

Bruce, R. A. (1984). Exercise, functional aerobic capacity, and aging—another viewpoint. In *Medicine and Science in Sports and Exercise* (Vol. 16, Issue 1, pp. 8–13). https://doi.org/10.1249/00005768-198401000-00004

Buckley, J. P., Sim, J., Eston, R. G., Hession, R., & Fox, R. (2004). Reliability and validity of measures taken during the Chester step test to predict aerobic power and to prescribe aerobic exercise. *British Journal of Sports Medicine*, *38*(2), 197–205. https://doi.org/10.1136/bjsm.2003.005389

Buckley, J., Sim, J., Eston, R. G., Hession, R., & Fox, R. (2004). Reliability and validity of measures taken during the Chester step test to predict aerobic power and to prescribe aerobic exercise. *British Journal of Sports Medicine*, *38*(2), 197–205. https://doi.org/10.1136/BJSM.2003.005389

Burr, J. F., Bredin, S. S. D., Faktor, M. D., & Warburton, D. E. R. (2011). The 6-minute walk test as a predictor of objectively measured aerobic fitness in healthy working-aged adults. *The Physician and Sportsmedicine*, *39*(2), 133–139. https://doi.org/10.3810/psm.2011.05.1904

Carvalho, L., Di Thommazo-Luporini, L., Aubertin-Leheudre, M., Bonjorno Junior, J., de Oliveira, C., Luporini, R. L., Mendes, R., Zangrando, K., Trimer, R., Arena, R., & Borghi-silva, A. (2015). Prediction of Cardiorespiratory Fitness by the Six-Minute Step Test and Its Association with Muscle Strength and Power in Sedentary Obese and Lean Young Women: A Cross-Sectional Study. *PloS One*, *10*(12). https://doi.org/10.1371/JOURNAL.PONE.0145960

Chéhère, B., Bougault, V., Gicquello, A., & Wallaert, B. (2016). Cardiorespiratory Response to Different Exercise Tests in Interstitial Lung Disease. *Medicine and Science in Sports and Exercise*, *48*(12), 2345–2352. https://doi.org/10.1249/MSS.0000000000001051

Cofre-Bolados, C., Rosales, W. D., & Espinoza-Salinas, A. (2018). Validation of the ST3X1 step test as an estimator of peak VO 2 in adults with cardiovascular risk factors. *Salud Uninorte*, *34*(3), 581–588. https://doi.org/10.14482/sun.34.3.616.10

Coquart, J. B., Grosbois, J.-M., Olivier, C., Bart, F., Castres, I., & Wallaert, B. (2016). Home-based neuromuscular electrical stimulation improves exercise tolerance and health-related quality of life in patients with COPD. *International Journal of Chronic Obstructive Pulmonary Disease*, *11*(1), 1189. https://doi.org/10.2147/COPD.S105049

Coquart, J. B., Lemaître, F., Castres, I., Saison, S., Bart, F., & Grosbois, J. M. (2015). Reproducibility and Sensitivity of the 6-Minute Stepper Test in Patients with COPD. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, *12*(5), 533–538. https://doi.org/10.3109/15412555.2014.974733

da Costa, J., Arcuri, J. F., Gonçalves, I., Davi, S., Pessoa, B., Jamami, M., & Di Lorenzo, V. A. P. (2014). Reproducibility of cadence-free 6-minute step test in subjects with COPD. *Respiratory Care*, *59*(4), 538–542. https://doi.org/10.4187/RESPCARE.02743

Dal Corso, S., Duarte, S. R., Neder, J. A., Malaguti, C., de Fuccio, M. B., de Castro Pereira, C. A., & Nery, L. E. (2007). A step test to assess exercise-related oxygen desaturation in interstitial lung disease. *European Respiratory Journal*, *29*(2), 330–336. https://doi.org/10.1183/09031936.00094006

Dal Corso, Simone, De Camargo, A. A., Izbicki, M., Malaguti, C., & Nery, L. E. (2013). A symptom-limited incremental step test determines maximum physiological responses in patients with chronic obstructive pulmonary disease. *Respiratory Medicine*, *107*(12), 1993–1999. https://doi.org/10.1016/j.rmed.2013.06.013

de Andrade, C. H. S., Cianci, R. G., Malaguti, C., & Corso, S. D. (2012). O uso de testes do degrau para a avaliação da capacidade de exercício em indivíduos saudáveis e pacientes com doenças pulmonares crônicas. *Jornal Brasileiro de Pneumologia*, *38*(1), 116–124. https://doi.org/10.1590/S1806-37132012000100016 de Andrade, C. H. S., De Camargo, A. A., de Castro, B., Malaguti, C., & Dal Corso, S. (2012). Comparison of cardiopulmonary responses during 2 incremental step tests in subjects with COPD. *Respiratory Care*, *57*(11), 1920–1926. https://doi.org/10.4187/RESPCARE.01742

De Blok, B. M. J., De Greef, M. H. G., Ten Hacken, N. H. T., Sprenger, S. R., Postema, K., & Wempe, J. B. (2006). The effects of a lifestyle physical activity counseling program with feedback of a pedometer during pulmonary rehabilitation in patients with COPD: A pilot study. *Patient Education and Counseling*, *61*(1), 48–55. https://doi.org/10.1016/j.pec.2005.02.005

de Camargo, A., Justino, T., de Andrade, C., Malaguti, C., & Dal Corso, S. (2011). Chester step test in patients with COPD: reliability and correlation with pulmonary function test results. *Respiratory Care*, *56*(7), 995–1001. https://doi.org/10.4187/RESPCARE.01047

Di Thommazo-Luporini, L., Pinheiro Carvalho, L., Luporini, R. L., Trimer, R., Falasco Pantoni, C. B., Catai, A. M., Arena, R., & Borghi-Silva, A. (2015). The Six-Minute Step Test as a predictor of cardiorespiratory fitness in obese women. *European Journal of Physical and Rehabilitation Medicine*, *51*(6), 793–802.

Eckel, R. H., Alberti, K. G. M. M., Grundy, S. M., & Zimmet, P. Z. (2010). The metabolic syndrome. *The Lancet*, *375*(9710), 181–183. https://doi.org/10.1016/S0140-6736(09)61794-3

Edvardsen, E., Hem, E., & Anderssen, S. A. (2014a). End criteria for reaching maximal oxygen uptake must be strict and adjusted to sex and age: A cross-sectional study. *PLoS ONE*, *9*(1), 18–20. https://doi.org/10.1371/journal.pone.0085276

Edvardsen, E., Hem, E., & Anderssen, S. A. (2014b). End Criteria for Reaching Maximal Oxygen Uptake Must Be Strict and Adjusted to Sex and Age: A Cross-Sectional Study. *PLoS ONE*, *9*(1). https://doi.org/10.1371/JOURNAL.PONE.0085276

Fabre, C., Chehere, B., Bart, F., Mucci, P., Wallaert, B., & Grosbois, J. M. (2017). Relationships between heart rate target determined in different exercise testing in COPD patients to prescribed with individualized exercise training. *International Journal of Chronic Obstructive Pulmonary Disease*, *12*, 1483. https://doi.org/10.2147/COPD.S129889

Falvey, J. R., Krafft, C., & Kornetti, D. (2020). The Essential Role of Home- and Community-Based Physical Therapists During the COVID-19 Pandemic. *Physical Therapy*, *100*(7), 1058–1061. https://doi.org/10.1093/PTJ/PZAA069

Geoffrey, J., Parsons, M., Mathieson, S., & Parsons, M. (2015). Home care: An opportunity for physiotherapy? *New Zealand Journal of Physiotherapy*, *43*(1). https://doi.org/10.15619/nzjp/43.1.04

Grosbois, J. M., Riquier, C., Chehere, B., Coquart, J., Béhal, H., Bart, F., Wallaert, B., & Chenivesse, C. (2016). Six-minute stepper test: A valid clinical exercise tolerance test for COPD patients. *International Journal of COPD*, *11*(1), 657–663. https://doi.org/10.2147/COPD.S98635

Hamilton, D., & Haennel, R. (2000). Validity and reliability of the 6-minute walk test in a cardiac rehabilitation population. *Journal of Cardiopulmonary Rehabilitation*, *20*(3), 156–164. https://doi.org/10.1097/00008483-200005000-00003

Hansen, D., Jacobs, N., Bex, S., D'Haene, G., Dendale, P., & Claes, N. (2011). Are fixed-rate step tests medically safe for assessing physical fitness? *European Journal of Applied Physiology*, *111*(10), 2593–2599. https://doi.org/10.1007/s00421-011-1886-3

Hong, S. H., Yang, H. I., Kim, D. Il, Gonzales, T. I., Brage, S., & Jeon, J. Y. (2019). Validation of submaximal step tests and the 6-min walk test for predicting maximal oxygen consumption in young and healthy participants. *International Journal of Environmental Research and Public Health*, *16*(23), 1–10. https://doi.org/10.3390/ijerph16234858

Howley, E. T., Bassett, D. R., & Welch, H. G. (1995). Criteria for maximal oxygen uptake: Review and commentary. *Medicine and Science in Sports and Exercise*, *27*(9), 1292–1301. https://doi.org/10.1249/00005768-199509000-00009

Huggett, D. L., Connelly, D. M., & Overend, T. J. (2005). Maximal aerobic capacity testing of older adults: A critical review. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *60*(1), 57–66. https://doi.org/10.1093/gerona/60.1.57

Iturian Barrón, A., Quintana Riera, S., & Reychler, G. (2021). The 3 Minute Step Test is a validated field test to evaluate the functional exercise capacity in children aged 6 to 12. *Respiratory Medicine and Research*, *80*. https://doi.org/10.1016/J.RESMER.2021.100833

Izquierdo, M. C., Lopes, S., Teixeira, M., Polónia, J., Alves, A. J., Mesquita-Bastos, J., & Ribeiro, F. (2019). The Chester step test is a valid tool to assess cardiorespiratory fitness in adults with hypertension: reducing the gap between clinical practice and fitness assessments. *Hypertension Research*, *42*(12), 2021–2024. https://doi.org/10.1038/s41440-019-0316-5

John Gormley, J. H. (Ed.). (2009). *Exercise Therapy: Prevention and Treatment of Disease*. John Wiley & Sons.

Jones, S., Tillin, T., Williams, S., Coady, E., Chaturvedi, N., & Hughes, A. D. (2017). Assessment of exercise capacity and oxygen consumption using a 6 min stepper test in older adults. *Frontiers in Physiology*, *8*(JUN), 2–8. https://doi.org/10.3389/fphys.2017.00408

José, A., & Corso, S. D. (2016). Step tests are safe for assessing functional capacity in patients hospitalized with acute lung diseases. *Journal of Cardiopulmonary Rehabilitation and Prevention*, *36*(1), 56–61. https://doi.org/10.1097/HCR.0000000000000149

Julio, U. F., Panissa, V. L. G., Shiroma, S. A., & Franchini, E. (2017). Effect of protocol manipulation for determining maximal aerobic power on a treadmill and cycle ergometer: A brief review. *Strength and Conditioning Journal*, *39*(5), 58–71. https://doi.org/10.1519/SSC.0000000000000332

Karloh, M., Corrêa, K., Martins, L., Araujo, C., Matte, D., & Mayer, A. F. (2013). Chester step test: assessment of functional capacity and magnitude of cardiorespiratory response in patients with COPD and healthy subjects. *Brazilian Journal of Physical Therapy*, *17*(3), 227–235. https://doi.org/10.1590/S1413-35552012005000087

Knight, E., Stuckey, M. I., & Petrella, R. J. (2014). Validation of the step test and exercise prescription tool for adults. *Canadian Journal of Diabetes*, *38*(3), 164–171. https://doi.org/10.1016/j.jcjd.2014.03.007

Laveneziana, P., Paolo, M. Di, & Palange, P. (2021). The clinical value of cardiopulmonary exercise testing in the modern era. *European Respiratory Review*, *30*(159), 1–9. https://doi.org/10.1183/16000617.0187-2020

Lemanska, A., Poole, K., Aning, J. J., Griffin, B. A., Manders, R., Saxton, J. M., Wainwright, J., & Faithfull, S. (2019). The Siconolfi step test: a valid and reliable assessment of cardiopulmonary fitness in older men with prostate cancer. *European Review of Aging and Physical Activity*, *16*(1), 1–10. https://doi.org/10.1186/s11556-018-0207-9

Magalhães, M. G. S., Teixeira, J. B., Santos, A. M. B., Clímaco, D. C. S., Silva, T. N. S., & de Lima, A. M. J. (2020). Construct validity and reproducibility of the six-minute step test in subjects with obstructive sleep apnea treated with continuous positive airway pressure. *Jornal Brasileiro de Pneumologia*, *46*(3), 1–7. https://doi.org/10.36416/1806-3756/e20180422

Manual., G. L. (2000). YMCA fitness testing and assessment. In *Champaign, IL: Human Kinetics.* (4th ed.). Published for the YMCA of the USA by Human Kinetics. https://www.worldcat.org/title/ymca-fitness-testing-and-assessment-manual/oclc/43798539#.XjLuLGgURP0.mendeley

Marrara, K., Marino, D., Jamami, M., Oliveira Junior, A., & Di Lorenzo, V. A. P. (2012). Responsiveness of the six-minute step test to a physical training program in patients with COPD. *Jornal Brasileiro de Pneumologia : Publicacao Oficial Da Sociedade Brasileira de Pneumologia e Tisilogia*, *38*(5), 579–587. https://doi.org/10.1590/S1806-37132012000500007

Mays, R. J., Boér, N. F., Mealey, L. M., Kim, K. H., & Goss, F. L. (2010). A comparison of practical assessment methods to determine treadmill, cycle and elliptical ergometer VO2peak. *Journal of Strength and Conditioning Research / National Strength & Conditioning Association*, *24*(5), 1325. https://doi.org/10.1519/JSC.0B013E3181C7C677

McGavin, C. R., Gupta, S. P., & McHardy, G. J. (1976). Twelve-minute walking test for assessing disability in chronic bronchitis. *British Medical Journal*, *1*(6013), 822. https://doi.org/10.1136/BMJ.1.6013.822

Mokkink, L. B., de Vet., H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., & Bouter, L. M. (2018). COSMIN risk of bias checklist for systematic reviews of patientreported outcome measures. *Quality of Life Research*, 1–37. www.cosmin.nl

Mokkink, L. B., Prinsen, C., Patrick, D., Alonso, J., Bouter, L., de Vet., H., & Terwee, C. (2018). *COSMIN manual for systematic reviews of PROMs COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) user manual*. www.cosmin.nl

Monteiro, D. P., Britto, R. R., Carvalho, M. L. V., Montemezzo, D., Parreira, V. F., & Pereira, D. A. G. (2014). Shuttle walking test como instrumento de avaliação da capacidade funcional: uma revisão da literatura. *Ciência & Saúde*, *7*(2), 92. https://doi.org/10.15448/1983-652x.2014.2.16580

Moore, J. P., Thom, J. M., Cooney, J. K., Maddison, P. J., Lemmey, A. B., Jones, J. G., Casanova, F., & Ahmad, Y. A. (2013). A Simple Step Test to Estimate Cardio-Respiratory Fitness Levels of Rheumatoid Arthritis Patients in a Clinical Setting. *International Journal of Rheumatology*, *2013*, 1–8.

Munari, A. B., Venâncio, R. S., Klein, S. R., Gulart, A. A., Silva, I. J. C. S., Sonza, A., Dal Lago, P., & Mayer, A. F. (2020). Physiological Responses to the 6-min Step Test in Patients with Chronic Obstructive Pulmonary Disease. *Journal of Cardiopulmonary Rehabilitation and Prevention*, *40*(1), 55–61. https://doi.org/10.1097/HCR.0000000000000469

Murphy, N., Bell, C., & Costello, R. (2005). Extending a home from hospital care programme for COPD exacerbations to include pulmonary rehabilitation. *Respiratory Medicine*, *99*(10), 1297–1302. https://doi.org/10.1016/J.RMED.2005.02.033

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*. https://doi.org/10.1136/BMJ.N71

Pessoa, B., Arcuri, J. F., Labadessa, I. G., Costa, J., Sentanin, A. C., & Di Lorenzo, V. A. P. (2014). Validity of the six-minute step test of free cadence in patients with chronic obstructive pulmonary disease. *Brazilian Journal of Physical Therapy*, *18*(3), 228–236. https://doi.org/10.1590/BJPT-RBF.2014.0041

Petrella, R. J., Koval, J. J., Cunningham, D. A., & Paterson, D. H. (2001). A Self-Paced Step Test to Predict Aerobic Fitness in Older Adults in the Primary Care Clinic. *JAGS*, *49*, 632–638.

Pichon, R., Couturaud, F., Mialon, P., Ber-Moy, C. Le, Péran, L., Lochon, C., Nowak, E., & Beaumont, M. (2016). Responsiveness and minimally important difference of the 6-minute stepper test in patients with chronic obstructive pulmonary disease. *Respiration*, *91*(5), 367–373. https://doi.org/10.1159/000446517

Prichard, R. A., Juul, M., Gazibarich, G., Davidson, P. M., Mason, C., Keogh, A. M., Macdonald, P. S., & Hayward, C. S. (2014). Six-minute walk distance predicts VO2 (max) in patients supported with continuous flow left ventricular assist devices. *International Journal of Artificial Organs*, *37*(7), 539–545. https://doi.org/10.5301/ijao.5000345

Prinsen, C. A. C., Mokkink, L. B., Bouter, L., Alonso, J., Patrick, D., de Vet., H., & Terwee, C. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *27*(5), 1147–1157. https://doi.org/10.1007/S11136-018-1798-3

Rajan, P. (2017). Community physiotherapy or community-based physiotherapy. *Health Promotion Perspectives*, *7*(2), 50–51. https://doi.org/10.15171/hpp.2017.10

Reed, J. L., Cotie, L. M., Cole, C. A., Harris, J., Moran, B., Scott, K., Terada, T., Buckley, J. P., & Pipe, A. L. (2020). Submaximal Exercise Testing in Cardiovascular Rehabilitation Settings (BEST Study). *Frontiers in Physiology*, *10*(January), 1–12. https://doi.org/10.3389/fphys.2019.01517

Reychler, G., Audag, N., Dewulf, S., Mestre, N. M., & Caty, G. (2018). Validation of 6 min step test and 4-m gait speed in children: A randomized cross-over study. *Gait and Posture*, *61*(August 2017), 19–24. https://doi.org/10.1016/j.gaitpost.2017.12.011

Ricci, P., Cabiddu, R., Jürgensen, S., André, L., de Oliveira, C., Di Thommazo-Luporini, L., Ortega, F., & Borghi-silva, A. (2019). Validation of the two-minute step test in obese with comorbibities and morbidly obese patients. *Brazilian Journal of Medical and Biological Research = Revista Brasileira de Pesquisas Medicas e Biologicas*, *52*(9). https://doi.org/10.1590/1414-431X20198402

Rikli, R. E., & Jones, C. J. (1998). The reliability and validity of a 6-minute walk test as a measure of physical endurance in older adults. *Journal of Aging and Physical Activity*, *6*(4), 363–375. https://doi.org/10.1123/JAPA.6.4.363

Rikli, R. E., & Jones, C. J. (1999). Development and Validation of a Functional Fitness Test for Community-Residing Older Adults. *Journal of Aging and Physical Activity*, *7*(2), 129–161. https://doi.org/10.1123/JAPA.7.2.129

Ritt, L. E. F., Darzé, E. S., Feitosa, G. F., Porto, J. S., Bastos, G., Albuquerque, R. B. L. de, Feitosa, C. M., Claro, T. C., Prado, E. F., Oliveira, Q. B. de, & Stein, R. (2021). O Teste do Degrau de Seis Minutos como Preditor de Capacidade Funcional de Acordo com o Consumo de Oxigênio de Pico em Pacientes Cardíacos. *Arquivos Brasileiros de Cardiologia*, *116*(5), 889–895. https://doi.org/10.36660/abc.20190624

RJ, P., JJ, K., DA, C., & DH, P. (2003). Can primary care doctors prescribe exercise to improve fitness? The Step Test Exercise Prescription (STEP) project. *American Journal of Preventive Medicine*, *24*(4), 316–322. https://doi.org/10.1016/S0749-3797(03)00022-9

Salbach, N. M., Brien, K. K., Brooks, D., Irvin, E., Martino, R., Takhar, P., Chan, S., & Howe, J. A. (2015). Reference values for standardized tests of walking speed and distance: A systematic review. *Gait & Posture*, *41*(2), 341–360. https://doi.org/10.1016/J.GAITPOST.2014.10.002

Santo, A. S., & Golding, L. A. (2003). Predicting maximum oxygen uptake from a modified 3-minute step test. *Research Quarterly for Exercise and Sport*, *74*(1), 110–115. https://doi.org/10.1080/02701367.2003.10609070

Sartor, F., Vernillo, G., De Morree, H. M., Bonomi, A. G., La Torre, A., Kubis, H. P., & Veicsteinas, A. (2013). Estimation of maximal oxygen uptake via submaximal exercise testing in sports, clinical, and home settings. *Sports Medicine*, *43*(9), 865–873. https://doi.org/10.1007/s40279-013-0068-3

SF, S., CE, G., TM, L., & RA, C. (1985). A simple, valid step test for estimating maximal oxygen uptake in epidemiologic studies. *American Journal of Epidemiology*, *121*(3), 382–390. https://doi.org/10.1093/OXFORDJOURNALS.AJE.A114010

Sheffield, L. T., Roitman, D., & Reeves, T. J. (1969). Submaximal exercise testing. *The Journal of the South Carolina Medical Association*, *65*(12), 782–807.

Starobin, D., Kramer, M. R., Yarmolovsky, A., Bendayan, D., Resenberg, I., Sulkes, J., & Fink, G. (2006). Assessment of functional capacity in patients with chronic obstructive pulmonary diesease: Correlation between cardiopulmonary exercise, 6 minute walk and 15 step exercise oximetry test. *Israel Medical Association Journal*, *8*(7), 460–463.

Stephenson, S., & Wiles, R. (2000). Advantages and disadvantages of the home setting for therapy: Views of patients and therapists. *British Journal of Occupational Therapy*, *63*(2), 59–64. https://doi.org/10.1177/030802260006300203

Stickland, M. K., Butcher, S. J., Marciniuk, D. D., & Bhutani, M. (2012). Assessing exercise limitation using cardiopulmonary exercise testing. *Pulmonary Medicine*. https://doi.org/10.1155/2012/824091

Strasser, B., & Burtscher, M. (2018). Survival of the fittest: VO2max, a key predictor of longevity? *Frontiers in Bioscience - Landmark*, *23*(8), 1505–1516. https://doi.org/10.2741/4657

Stuckey, M., Knight, E., & Petrella, R. J. (2012). The Step Test and Exercise Prescription Tool in Primary Care: A Critical Review. *Critical Reviews™ in Physical and Rehabilitation Medicine*, *24*(1–2), 109–123. https://doi.org/10.1615/CRITREVPHYSREHABILMED.2013006823

Swinburn, C. R., Wakefield, J. M., & Jones, P. W. (1985). Performance, ventilation, and oxygen consumption in three different types of exercise test in patients with chronic obstructive lung disease. *Thorax*, *40*, 581–586.

Sykes, K., & Roberts, A. (2004). The Chester step test—a simple yet effective tool for the prediction of aerobic capacity. *Physiotherapy*, *90*(4), 183–188. https://doi.org/10.1016/J.PHYSIO.2004.03.008

Taylor, N. F., Dodd, K. J., Shields, N., & Bruder, A. (2007). Therapeutic exercise in physiotherapy practice is beneficial: A summary of systematic reviews 2002-2005. *Australian Journal of Physiotherapy*, *53*(1), 7–16. https://doi.org/10.1016/S0004-9514(07)70057-0

Teren, A., Zachariae, S., Beutner, F., Ubrich, R., Sandri, M., Engel, C., Löffler, M., & Gielen, S. (2016). Incremental value of Veterans Specific Activity Questionnaire and the YMCA-step test for the assessment of cardiorespiratory fitness in population-based studies. *European Journal of Preventive Cardiology*, *23*(11), 1221–1227. https://doi.org/10.1177/2047487315621844

Travensolo, C. de F., Arcuri, J. F., & Polito, M. D. (2020). Validity and reliability of the 6-min step test in individuals with coronary artery disease. *Physiotherapy Research International*, *25*(1), 1–7. https://doi.org/10.1002/pri.1810

Travensolo, C., & Polito, M. (2014). Testes de Degrau para Avaliação da Capacidade de Exercício em Pacientes com Doenças Cardíacas: Resvisão da Literatura. *Revista Brasileira de Cardiologia*, *27*(6), 445–453.

Tveter, A. T., Dagfinrud, H., Moseng, T., & Holm, I. (2014). Measuring health-related physical fitness in physiotherapy practice: Reliability, validity, and feasibility of clinical field tests and a patient-reported measure. *Journal of Orthopaedic and Sports Physical Therapy*, *44*(3), 206–216. https://doi.org/10.2519/jospt.2014.5042

Vanhelst, J., Fardy, P. S., Salleron, J., & Béghin, L. (2013). The six-minute walk test in obese youth: Reproducibility, validity, and prediction equation to assess aerobic power. *Disability and Rehabilitation*, *35*(6), 479–482. https://doi.org/10.3109/09638288.2012.699581

Vieira, E. B., Ota-Arakaki, J. S., Dal Corso, S., Ivanaga, I., Fonseca, A. X. C., Oliveira, R. K. F., Rodrigues-Júnior, J. C., Ferreira, E. V. M., Nery, L. E., & Ramos, R. P. (2020). Incremental step test in patients with pulmonary hypertension. *Respiratory Physiology and Neurobiology*, *271*(September 2019). https://doi.org/10.1016/j.resp.2019.103307

Vilarinho, R., Caneiras, C., & Montes, A. M. (2021). Measurement properties of step tests for exercise capacity in COPD: A systematic review. *Clinical Rehabilitation*, *35*(4), 578–588. https://doi.org/10.1177/0269215520968054

Vilarinho, R., Mendes, A., Gomes, M., Ferreira, R., Costa, F., Machado, M., Neves, M., Caneiras, C., & Montes. (2021). Adapted Chester Step Test Can Have Maximal Response Characteristics for the Assessment of Exercise Capacity in Young Women. *Healthcare (Basel, Switzerland)*, *9*(3). https://doi.org/10.3390/HEALTHCARE9030308

WCPT. (2011). *Policy statement: Description of physical therapy | World Physiotherapy*. https://world.physio/policy/ps-descriptionPT

Winett, R. A., & Carpinelli, R. N. (2000). Examining the validity of exercise guidelines for the prevention of morbidity and all-cause mortality. *Annals of Behavioral Medicine*, *22*(3), 237–245. https://doi.org/10.1007/BF02895119

Wise, R. A., & Brown, C. D. (2005). Minimal clinically important differences in the six-minute walk test and the incremental shuttle walking test. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, *2*(1), 125–129. https://doi.org/10.1081/COPD-200050527

Wyndham, C. H. (1967). Submaximal tests for estimating maximum oxygen intake. *Canadian Medical Association Journal*, *96*(12), 736–745.

Appendix 1 – Key terms and data bases.

PubMed + Medline	((("step test") OR ("stepper test")) AND ("exercise testing")) OR ((("step test") OR ("stepper test")) AND (aerobic capacity))
Cochrane Central	((("step test") OR ("stepper test")) AND ("exercise testing")) OR ((("step test") OR ("stepper test")) AND (aerobic capacity))
PEDRo	step test OR stepper test
LILACs	step test OR stepper test
Web of Science	"step test" OR "stepper test" AND "exercise testing" AND "cardiopulmonary testing"

Appendix 2 – Study Design: clinical group, sample size and characteristics, protocol and outcomes.

Reference	Study Design	Step Test	Clinical Group	n	Protocol	Outcomes
(De Blok et al., 2006)	Pilot study	2MST	COPD	Experimental: 10 (5 males); 65,7±10,4 years Control: 11 (4 males); 62,5±12,3	9 week RP; Experimental group with lifestyle activity counselling program. Both groups performed 2MST before and after RP	Primary outcomes – daily physical activity (steps/day); Secondary outcomes – physical fitness (2MST – number of steps), HRQoL
(Ricci et al., 2019)	Observational study	2MST	Obesity and morbid obesity	31 (19,0% males) 37,3±3,6 years	Visit 1: anamnese, anthropometry, CPET on cycle ergometer Visit 2: 2MST	Primary outcome - number of steps in 2MST and VO_{2max} in CPET
(Beaumont et al., 2019)	Cross-over observational	3MST	50 (33 males) COPD patients	50 (33 males) 62,3±7,9 years	Random order Same day: 3MST; 6MWT	Primary outcome - number of repetitions during the 3MST and the 6MWT distance; Secondary Outcomes - HR, SpO_2, dyspnea and leg fatigue (before and after) and HR and SpO_2 evolution of during both tests; reliability of the 3MStepT.
(Borel et al., 2016)	randomized, double- blind, placebo-controlled, crossover study	3MST	COPD	40 (33 males) 69±7 years	All with bronchodilator Visit 1 – clinical assessment; Visit 2 – CPET and pulmonary function testing Visit 3 and 4 - 3MST at 14,16,20,24 steps/minute-	Primary outcome – dyspnea (mBorg), physiological responses, VO2
(Hong et al., 2019)	Cross-section observational	3MST	Healthy	73 (37 males) 30,8±9,3 years	Group A/Group B (random order on visit 2 and 3) Visit 1: Anthropometric measures, clinical evaluation and CPET on treadmill Visit 2: 3MST with 20,3/30cm box Visit 3: 3MST with 20,3/30 cm box	Primary outcomes - VO_{2max}, anthropometric measurements, and outcome measures from the step tests and 6MWT
(Iturian Barrón et al., 2021)	Randomized Cross-section observational	3MST	Healthy	30 children (15 females) 8,6±1,8 years	Visit 1: 6MWT/3MST (50% each) Visit 2: 6MWT/3MST (50% each) Visit 3: 3MST	Primary outcome - number of steps, distance in 6MWT; Secondary outcomes - cardiopulmonary responses.

(Murphy et al., 2005)	Prospective, controlled and randomised study	3MST	COPD	Experimental group: 13 (7 males); 67,7±9.7 years Control group: 13 (10 males); 65,7±11 years	Visit 1 – 3MST, Shuttle Walk Test, clinical assessment. 6 weeks RP Visit 2 - 3MST, Shuttle Walk Test, clinical assessment.	Primary outcome – level reached on Shuttle Walk Test, number of steps, HR, dyspnea (mBorg)
(Cofre-Bolados et al., 2018)	Cross-sectional, correlational	3MSTx1	Cardiovascular risk factors	111 (75,67% males) 49,81±11,16 years	Visit 1: CPET on cycle ergometer Visit 2: 3MSTx1	Primary outcomes - number of steps, VO_2
(Bonnevie et al., 2017)	Retrospective Study	6MSrT	COPD	24 (6 females) 61.5±8.7 years	Two 6MSrT on the same day; CPET Pulmonary function assessment	Primary outcome – development of regression equation to predict the HR at the ventilatory threshold; Secondary outcome – correlation between step count and ventilatory threshold from CPET.
(Bonnevie et al., 2019)	Retrospective Study	6MSrT	COPD	35 (10 females) 61±9 years	Retrospective study: Assessed CPET on treadmill, 1RM, 6MSrT, 6MWT	Primary outcomes – number of steps, distance walked, 1RM, W_{peak}
(Coquart et al., 2015)	Prospective Observational Study	6MSrT	COPD	35 (14 females) 60.8±8.9 years	Visit 1 – 2 6MSrT 6 weeks RP Visit 2 – 2 2 6MSrT	Primary outcomes – number of steps, physiological responses (HR, dyspnea Borg)
(Coquart et al., 2016)	Retrospective, observational study	6MSrT	COPD	188	Visit 1 – clinical assessment, 6MSrT Group NMES – 6 weeks of home program with NMES Group UEPE – 8 weeks of RP program. Visit 2 – clinical assessment, 6MSrT	Primary outcome – number of steps, phy
(Fabre et al., 2017)	Monocentric retrospective study	6MSrT	COPD	50 (7 females) 60.1±8.5 years	Visit 1 – spirometry Visit 2 – CPET Visit 3 – 6MSrT, 6MWT	Primary outcome – heart rate from the three assessments

(Pichon et al., 2016)	Observational Study	6MSrT	COPD	62 (37 males) 61.8 ± 9 years	Visit 1 – 6MSrT, 6MWT, clinical assessment and dyspnea, leg discomfort, heart rate 5 week RP Visit 1 – 6MSrT, 6MWT, clinical assessment and dyspnea, leg discomfort, heart rate	Primary outcome – difference in the number of steps before and after RP Secondary outcome – minimal important difference for 6MSrT, SpO_2, dyspnea, leg discomfort, heart rate for 6MSrT
(Arcuri et al., 2016)	Cross-sectional, correlational	6MST	Healthy	91 (49 males) 39±17 years	**Visit 1:** general assessment, spirometry, 2 6MST/2 6MWT **Visit 2:** body composition analysis, 2 6MST/2 6MWT	**Primary outcomes** - number of steps (comparing both 6MST), distance walked (compared with 6MST).
(Pessoa et al., 2014)	Observational Study	6MST	COPD	32 (24 male) 69,0±10,0 years	Random order Visit 1 – clinical and risk factor assessment; 6MST/6MWT Visit 2 – BODE; 6MST/6MWT	Primary outcome – number of steps, walked distance, BODE index score
(Di Thommazo-Luporini et al., 2015)	Cross-section observational	6MST	Obesity	56 women 35±7 years	**Visit 1;** Clinical and physiotherapeutic evaluations **Visit 2:** CPET on treadmill **Visit 3:** 6MST	**Primary outcome** - data from CPET and 6MST (VO_{2peak} and HR)
(Grosbois et al., 2016)	Monocentric retrospective cohort study	6MST	COPD	91 (76 males) 60,3±9,3 years	**Retrospective** 6MST, 6MWT and CPET on cycle ergometer data from before PR 6MST and 6MWT data from after PR	**Primary outcome** - data from 6MST, 6MWT, and CPET **Secondary outcome** - data from before/after PR
(da Costa et al., 2014)	Observational Study	6MST	COPD	32 68,3±10,3 years	Visit 1 – clinical assessment; 2 6MST Visit 2- 6MST	Primary outcome - Number of steps
(Marrara et al., 2012)	Controlled, prospective randomized study	6MST	COPD	36 males 69,4±8,6 years	Pre RP and Post RP Visit 1 - clinical assessmental Visit 2 – CPET Visit 3 – 6MST Visit 4 – 6MWT	**Primary outcomes** - VO_{2max} in CPET; number of steps, 6MWT distance, physiological responses.
(Carvalho et al., 2015)	Prospective Cohort	6MST	Healthy and obese	23 females (13 lean (LN) 32±5 years; 18 obese (OB)35±5 years.)	**Visit 1:** Clinical screening and assessment of anthropometrics and pulmonary function **Visit 2:** Blood analysis	**Primary outcome** - VO_{2peak} on CPET; number of steps in 6MST; muscle mass and function.

					Visit 3: CPET on treadmill **Visit 4 and 5:** 6MST **Visit 6:** Isometricand Isokinetic Strength Measures	
(Magalhães et al., 2020)	Cross-section observational	6MST	OSA	48 patients 54,5 years	Random order **Visit 1:** 6MWT; 6MST **Visit 2:** 6MWT/6MST	**Primary outcome** – number of steps during the 6MST and the 6MWT distance;
(Munari et al., 2020)	Cross-sectional observational study	6MST	COPD	36 (29 males) 67±7 years	Visit 1 – Pulmonary function, dyspnea, health status, and functional status Visit 2 – physiological variables during the 6MST	Primary outcome – number of steps, physiological variables
(Reychler et al., 2018)	Randomized cross-over study	6MST	Healthy	60 children (33 females) 8,0±1,3 years	Random order **Visit 1:** 6MWT/6MST/4MGS **Visit 2:** 6MWT/6MST/4MGS **Visit 3:** 6MWT/6MST/4MGS	**Primary outcome** – results of the three tests, and physiological responses (HR, SpO_2)
(Ritt et al., 2021)	Cross-section observational	6MST	Coronary artery disease	171 (121 males) 60±14 years	**Visit 1:** CPET on treadmill/6MST **Visit 2:** CPET on treadmill/6MST	**Primary outcomes** – number of steps vs. VO_{2peak}
(S. Dal Corso et al., 2007)	Cross-section observational	6MST	Interstitial lung disease	31 (12 females) 58±10 years	**Visit 1:** CPET max on cycle ergometer; 6MST **Visit 2:** constant rate CPET (submaximal) on cycle ergometer; 6MST	**Primary outcomes** –number of steps; step $VO_{2\,peak}$; VO_{2peak} on CPET, peak HR_{peak} on CPET, RER on CPET, rating of perceived exertion
(C. de F. Travensolo et al., 2020)	Cross-sectional, randomized, and crossover	6MST	Coronary artery disease	35 (23 males) 65.8±9.6 years	**Visit 1:** 6MST/6MWT **Visit 2:** 6MST/6MWT	**Primary outcomes** – number of steps, distance walked in 6MWT.
(Vilarinho, Mendes, et al., 2021)	Cross-section observational	aCST	Healthy sedentary	25 females 20.3±1.5 years	**Visit 1: aCST**; ISWT **Visit 2:** aCST; ISWT	**Primary outcomes** – Number of steps vs total distance **Secondary outcomes** – on the ACST: maximum work rate reached, total test time; on the ISWT: gait speed reached, total test time.
(de Camargo et al., 2011)	Cross-section observational	CST	COPD	32 (3 female) 69,9±9,0 years	Random order **Visit 1:** 2 6MWT/CST/CPET on cycle ergometer	**Primary outcome** – number of steps on CST; **Secondary outcomes** – HR, dyspnea and leg fatigue by Borg, SpO_2 plus walked distance in 6MWT and VO_{2peak} in CPET.

					Visit 2: 2 6MWT/CST/CPET on cycle ergometer	
(José & Corso, 2016)	Cross-section observational	CST MIST	Acute lung disease	97 individuals (77 patients (36 males); 56,3 (36-68) years/20 control; 52 (41-66) years)	**Visit 1:** CST/MIST **Visit 2:** 6MWT	**Primary outcome** – number of steps, Distance in 6MWT; **Secondary outcomes** – pulmonary function, dyspnea, time of hospitalization.
(J. Buckley et al., 2004)	Cross-section observational	CST	Healthy	13 (7 males) 22,4±4,6 years	**Visit 1:** CST **Visit 2:** CST **Visit 3:** CPET on treadmill	**Primary outcome** – reliability of the CST (predicted VO_{2max}, HR, RPE, and measured VO_{2max}); **Secondary outcome** – validity of the CST (predicted VO_{2max} vs measured VO_{2max} in CPET, $\%HR_{max}$ with HR_{max} in CPET; RPE)
(M et al., 2013)	Cross-section observational	CST	Healthy and COPD	Control group (healthy) – 10 (7 females) 63±7 years COPD group – 10 (7 females) 64±10 years	**Visit 1:** Anamnesis, anthropometric and lung function evaluation and ISWT to all individuals; modified Baecke questionnaire to the control group **Visit 2:** 6MWT to all individuals and London Chest Activity of Daily Living Scale (LCADL) to COPD group **Visit 3:** CST to all individuals	**Primary outcomes** – Performance in ISTW, 6MWT and CST and the HR, SpO_2 and dyspnea sensation measured in CST.
(Reed et al., 2020)	Observational prospective study	CST	Cardiovascular disease	47 (77% males) 57±11 years	Baseline CPET + MBTT (Modified Bruce Treadmill Test) + CST (Chester) + ARCT (Astrand-rhynming) 12 week rehabilitation program Follow-up Phase CPET + MBTT (Modified Bruce Treadmill Test) + CST (Chester) + ARCT (Astrand-rhynming)	**Primary outcome** – data from functional tests and CPET
(Simone Dal Corso et al., 2013)	Cross-section observational	IST	COPD	34 males 67±9 years	**Visit 1:** CPET on cycle ergometer; IST-1 **Visit 2:** IST-2	**Primary outcomes** – VO_2 and number of steps

(Vieira et al., 2020)	Prospective cross-sectional	IST	Pulmonary hypertension	20 (13 females) 41±15 years	Random order **Visit 1:** IST/CPET on cycle ergometer **Visit 2:** IST/CPET on cycle ergometer	ND
(Swinburn et al., 1985)	Observational Study	PST	COPD	17 (6 males) 66 (49-73) years	Same day – 12MWT, PST, CPET	**Primary outcome – number of steps, distance walked on 12MWT, VO_{2max},**
(Hansen et al., 2011)	Cross-section observational	SFRST	Healthy	113 (60 males) 45±13 years	**Visit 1:** CPET on cycle ergometer **Visit 2:** Fixed-Rate Step Test	**Primary outcomes** – maximal physiological response (HR and VO_2) between CPET and step test **Secondary outcomes** – %CPET VO_{2max} and subject characteristics; Relationship between %CPET VO_{2max} independent variable) and age, gender, body height, body mass index, VO_{2max}, VT, and high-intensity physical activity (dependent variables).
(Lemanska et al., 2019)	Cross-section observational	SST	Prostate cancer	83 males 68,2±7,4 years	**Visit 1:** CPET on cycle ergometer and SST **Visit 2:** SST	**Primary outcomes** – number of steps, VO_{2peak}
(Moore et al., 2013)	Prospective validation	SST	Rheumatoid arthritis	30 (24 females) 53,4±10,4 years	**Visit 1:** Siconolfi step test **Visit 2:** Siconolfi step test + CPET on cycle ergometer	**Primary outcomes** – CPET: VO_{2max} and ventilatory variables from gas analysis; On the step test: number of steps; and ventilatory variables from gas analysis.
(Knight et al., 2014)	Cross-sectional, correlational	STEP	Healthy	40 (20 males) 43±14 years	Random order **Same day** STEP; CPET on treadmill	**Primary outcomes** – predicted VO_{2max}; measured VO_{2max}.
(Beutner et al., 2015b)	Cross-section observational	YMCA	Healthy	97 (55 males) 55,7±14,2 years	**Same day** YMCA Step Test; CPET on treadmill	**Primary outcome** - VO_2 on CPET, YMCA
(Teren et al., 2016)	Prospective Cohort	YMCA	Overweight	105 (53 males) 60 (46-67) years	**One visit sequence:** 1. YMCA-Step Test > 2. VSAQ > 3. CPET on treadmill	**Primary outcome** - VO_2 in step test, compared with VO_{2peak} in CPET and scores on VSAQ (The Veterenas Specific Activity Questionnaire).

2MST - 2-Minute Step Test; 3MST - 3-Minute Step Test; 6MST - 6-min Step Test; 6MSrT - 6-minute Stepper Test; 12MWT – 12 Minute Walking Test; aCST - Adapted Chester Step Test; CST - Chester Step Test; IST - Incremental Step Test; mIST - Modified Incremental Step Test; PST – Fixed Rate Paced Step Test; SST - Siconolfi Step Test; SFRST - Standardized Fixed-Rate Step Test; ST3x1 - Step Test 3x1; STEP - Step Test and Exercise Prescription (STEP); YMCA - YMCA-Step Test; +1mHBC - +1-Minute Heart Beat Count; HRQoL – Health Related Quality of Life Questionnaire; 4MGS – 4-Meter Gait Speed; CPET – cardiopulmonary exercise testing; NMES – Neuromuscular Electrical Stimulation; 1RM – one repetition maximum; RP – rehabilitation program; VO_{2max} – maximal oxygen uptake; HR_{max}– maximal heart rate; %HR_{max} – percentage of maximal heart rate; RER - respiratory exchange ratio; Borg – Borg Exertion Scale; mBorg – Modified Borg Exertion Scale; VAS –

Visual Analog Scale; $\%VO_{2max}$ – percentage of VO_{2max}; VO_{2peak} – peak of maximal oxygen uptake; OSA – Obstructive Sleep Apnea; mL – millilitres; kg – kilogram; min – minute; bpm – beats per minute; ND – Not described;

Appendix 3 – Outcomes – measurement properties and physiological responses.

Reference	Measurement Properties			Physiological Responses of the Step Test
	Validity	*Reliability*	*Responsiveness*	
(de Camargo et al., 2011)	Construct Validity number of steps vs. walked distance; r = 0.60, p<0.001 Criterion Validity Estimated VO_2 on CST vs. direct measured VO_2 at CPET. r = 0.69, p<0.02	number of steps on test 1 vs. Number of steps on test 2 ICC = 0.99 (0.97–0.99)		$\%HR_{max}$: CST 1 - 80±10; CST 2 - 80±13; Dyspnea mBorg: CST 1 - 4 (1-10); CST 2 - 4 (0-9); Leg fatigue mBorg: CST 1 - 4 (2-10); CST 2 - 5 (0-10).
(Arcuri et al., 2016)	Construct Validity number of steps vs. distance walked on 6MWT r = 0.72; p<0.05	Step test 1 vs. Step test 2 ICC = 0.9; 95% CI: 0.85–0.93.		ND
(Beaumont et al., 2019)	Construct Validity number of steps vs. distance walked r=0.780; p<0.001	Step test 1 vs. Step test 2 ICC=0.964 (95%CI=0.936–0.979)		HR (bpm) - 112,9±13,4; Dyspnea (mBorg) - 5,0±2,2; Lower limb fatigue (mBorg) - 3,4±2,4.
(Bonnevie et al., 2017)	Criterion Validity Number of steps vs. Ventilatory threshold from CPET R=0,63; p<0.01			HR_{peak} = 120,6 ±.15 bpm

(Bonnevie et al., 2019)	Criterion Validity Number of steps vs. W_{peak} r=0.44, P,0.01 Construct Validity Number of steps vs. 1RM r=0.41, P=0.01			ND
(Borel et al., 2016)			To bronchodilator induced relief on exertional dyspnea (significantly decreased dyspnea at the end of the third minute of exercise at 14 steps/min (by 0.6±1.0 Borg 0–10 scale units, P,0.01) and 16 steps/min (by 0.7±1.3 Borg 0–10 scale units, P,0.01))	$\%HR_{max}$ – 83 – 91% (experimental e control group) Dyspnea (mBorg) – 2,3 – 5,9 Leg discomfort (mBorg) – 2,5 – 4,9
(Pessoa et al., 2014)	Construct Validity Number of steps vs. Walked distance r=0,764; P=0,000			ND
(Cofre-Bolados et al., 2018)	Criterion Validity number of steps/predicted VO_{2peak} vs. measured VO_{2peak} r = 0.81, p<0.05			Not described

(Coquart et al., 2015)		Number of steps 6MSrT 1 vs number of steps 6MSrT Pre RP – $r = 0.92$; $p < 0.01$; ICC = 0.919; Cronbach's α = 0.952 Post RP – $r = 0.95$; $p < 0.01$; ICC = 0.945; Cronbach's α = 0.972	6MSrT (number of steps) to RP Number of strokes did not differ from pre to post RP $P < 0.01$; $r = 0.32$	Pre RP $\%HR_{max}$ = 72,4±3,4 Dyspnea (mBorg) = 3,8±0,6 Leg fatigue (mBorg) = 4,1±0,7 Post RP $\%HR_{max}$ = 70,4±3,3 Dyspnea (mBorg) = 3,8±0,6 Leg fatigue (mBorg) = 3,7±0,5
(Coquart et al., 2016)			To PR or NMES 6MSrT performances were improved after the intervention period in both the groups, $p<0.005$.	ND
(De Blok et al., 2006)			Responsiveness to RP ('2-min step test' with a large between-groups effect size (>0.80))	ND
(Di Thommazo-Luporini et al., 2015)	Criterion Validity 6MST VO_{2peak} vs. CPET $VO_{2\,peak}$ $r=0.56$; $P<0.001$			VO_{2peak} – 17,8±3,6 mL/kg/min RER – 1,03±0,09 $\%HR_{max}$ – 87±9 bpm Dispnea mBorg – 3 (0;10) Leg fatigue mBorg – 3(0;100)
(Beutner et al., 2015b)	Criterion validity step test VO_{2max} vs. CPET VO_{2peak}. $r=0.862$, $p<0.001$			RER_{peak} – 1,02±0,14

(Fabre et al., 2017)	Criterion Validity Heart rate from 6MSrT vs. heart rate from ventilatory threshold CPET r=0.48 Construct Validity Heart rate from 6MSrT vs. heart rate from 6MWT r=0.68 and r=0.62			HR_{peak} – 118 (83–152) bpm
(Hansen et al., 2011)	Criterion validity step test VO_{2max} vs. CPET VO_{2peak}. r = 0.54, p<0.01			%pHR_{max} = 86 ± 12; RPE Borg = 13 ± 2.
(Hong et al., 2019)	Construct Validity number of steps vs. distance walked on 6MWT Criterion Validity number of steps vs. VO_{2max} $3MST_{20}$ – r = 0.866 p < 0.05,$3MST_{30}$ – r = 0.859, p < 0.05			HR_{peak} with 20,3cm box - 116,7±15,0 bpm HR_{peak} with 30,0cm box - 137,5±17,8 bpm
(Iturian Barrón et al., 2021)	Construct Validity number of steps vs. distance walked in 6MWT 3MST1: r = 0.833; p<0.001 and 3MST2: r = 0.868; p<0.001	Step test 1 vs. Step test 2 ICC = 0.944-95%ci = 0.771;0.979; P<0.001.		HR_{peak} – 147±2,3 bpm Fatigue mBorg - 4,13±2,2
(Grosbois et al., 2016)	Construct Validity number of steps vs. distance walked r=0.56; P<0.0001 Criterion Validity number of steps vs. VO2max r=0.39; P<0.005		Responsiveness: to PR – improvements on 6MST, r=0.34; P=0.03	Not described

(da Costa et al., 2014)		Number of steps 1 vs number of steps 2 vs. number of steps 3 ICC = 0.80 – 0.94; mean error of 5.7 steps and limits of agreement between −7 and 18 steps		ND
(José & Corso, 2016)	**Construct Validity** number of steps on each step test vs. distance walked in 6MWT (Patients group) CST - r = 0.59, p<0.001; MIST - r = 0.64, p<0.001			**Patients (CST/MIST)** $\%HR_{max}$ - 76±11/76±12 mBorg dyspnea - 4 (3-7)/4 (3-7) mBorg leg fatigue - 5 (3-7)/5 (3-7) **Controls (CST/MIST)** $\%HR_{max}$ - 85±13/85±13 mBorg dyspnea - 2 (1-3)/2 (1-3) mBorg leg fatigue - 2 (1-3)/2 (1-4)
(J. Buckley et al., 2004)	**Criterion Validity** VO_{2peak} CPET vs. VO_2 predicted by step test bias±95% LoA −22.8 ml/kg/min (p=0.006) (trial 1)	Step test 1 vs Step Test 2 bias±95% LoA 0.8 (3.7) ml/kg/min (p>0.05)		**RPE (Borg)** - 14,2±2,0 **$\%HR_{max}$** - 81,1±11,2 **$\%VO_{2max}$** - 65,6±9,3
(Knight et al., 2014)	**Criterion Validity** predicted VO_{2peak} vs. measured VO_{2peak} r=0.78, p<0.001			HR_{peak} - 174±21 bpm
(Marrara et al., 2012)			Responsiveness to RP (effect size of approximately 11 steps on physical training group)	**Pre RP (Experimental/Control)** **mBorg – 1/0.5 (0,5–5,0)** **Post RP (Experimental/Control)** **mBorg – 0 (0-2)/0 (0-5)**
(Lemanska et al., 2019)	**Criterion Validity** number of steps vs. VO_{2peak} r=0.69, p<0.001, 95%CI 0.54 to 0.80	Step test 1 vs. Step test 2 ICC = 0.81; p < 0.001, 95% CI 0.70 to 0.89.		HR_{max} - 141,6±20,7 bpm

(Carvalho et al., 2015)	**Criterion Validity** VO_{2peak} CPET vs. $VO_{2predicted}$ by step test r = 0.86, p<0.05			LN VO_{2peak} (ml/kg/min) - 24,9±3,7 RER - 1,04±0,09 $\%HR_{max}$ - 92±8 mBorg dyspnea - 3 (0,5-8) mBorg leg fatigue - 3 (0-7) OB VO_{2peak} (ml/kg/min) - 17,6±3,1 RER - 1,07±0,10 $\%HR_{max}$ - 90±9 mBorg dyspnea - 3,5 (0-10) mBorg leg fatigue - 2 (0-10)
(Karloh et al., 2013)	**Construct Validity** number of steps CST vs. total distance in 6MWT/level on ISWT r=0.76, p<0.02. (CST vs. 6MWT); r=0.76, p<0.01 (CST vs. ISWT)			**% HRmax in CST:** Control group - 83,5±14,9%; COPD group - 70,5±10,5%; **ΔBorg in CST:** Control group - 2,6±2,1%; COPD group - 2,6±2,7%.
(Magalhães et al., 2020)	**Construct Validity** number of steps vs. distance walked r = 0.520; p < 0.001	Step test 1 vs Step Test 2 0.976 (0.957-0.986, p<0.001.		$\%HR_{max}$ - 74,2±11,8; Leg fatigue (mBorg) - 5,1±1,9.
(Moore et al., 2013)	**Criterion validity** mean values for VO_{2max} estimated from the second step test vs. directly measured VO_{2peak}. r = 0.79, LoA ±5.7mL·kg-1·min-1, (95% CI 0.55 to 0.91)	Step test on visit 1 vs. Step test on visit 2 ICC = 0.97, LoA ±2.2mL·kg-1·min-1, (95% CI of 0.94 to 0.99), r = 0.97 (95% CI 0.93 to 0.99)		$\%HR_{max}$ on visit 1 - 73±8%; % HR $_{max}$ on visit 2 - 75±8%.
(Munari et al., 2020)		Number of step 1 vs. number of step 2 ICC = 0.98, P < .001 SEM of 4.27 and an MDC of 11.8		Described VO_{2max} plateau HR_{peak} – 119 ± 12.3 ($\%HR_{max}$ ≈ 81.5%)

(Murphy et al., 2005)			Responsiveness to a RP; pre-post: Experimental group increased from 119 to 163 steps (both P<0:001)	HR_{peak} Experimental – pre: 112±13; post: 116±13 Control – pre: 118,8±15; post: 113±10 Dyspnea (mBorg) Experimental – pre: 5,5±2,55; post: 4,9±1,76 Control: pre – 5,1±3,24; post: 5,3±3,6
(Ricci et al., 2019)	Criterion Validity number of steps vs. VO_{2max} r=0.70, P<0.001 (VO_{2peak} from CPET and 2MST); r=0.55; P=0.01 (VO_{2peak} vs. step cicles)			VO_2 (mL/kg/min) - 12,5 (11,8-13,2) RER - 1,01 (0,96-1,06) $\%HR_{max}$ - 74,3 (70,9-77,7) mDyspnea Borg - 2 (0-7) mLeg fatigue - 0 (0-6)
(Pichon et al., 2016)	Construct Validity Number of steps vs. Walked distance Pre – r = 0.72; p < 0.0001 Post – r = 0.66; p < 0.0001		To RP. Number of steps raised 22.5 steps; 95% CI 13.8–31.3; p < 0.0001. MID was estimated to be around 20 steps SEM was 17.6 and 23.5 steps	Pre RP HR_{peak} – 117.9 ± 15.8 bpm Leg discomfort (VAS) – 4.9 ± 2.1 cm Dyspnea (mBorg) - 5.4 ± 2.0 Post RP HR_{peak} – 118.9 ± 17 bpm Leg discomfort (VAS) – 4.5 ± 2.0 cm Dyspnea (mBorg) - 4.9 ± 2.1
(Reed et al., 2020)	Criterion Validity number of steps vs. VO_2 peak r = 0.693, p < 0.001		Responsiveness to cardiovascular rehabilitation program: Δ1.9 ± 5.0 mL.kg-1.min-1, p = 0.300; n = 29 in $VO_{2max.}$	Not described
(Reychler et al., 2018)	Construct Validity number of steps of 6MST vs. distance walked on 6MWT r = 0.320; p = 0.013			HR_{peak} - 1129±25 bpm Fatigue VAS - 5,40 (0,5-10,0mm)
(Ritt et al., 2021)	Criterion Validity number of steps vs. VO_{2peak}			RER - 1,12±0,8

	r=0.69; p <0.001			
(S. Dal Corso et al., 2007)	**Criterion validity** mean values for VO_{2max} from the step test vs. $CPET_{max}$ measured VO_{2peak}. r = 0.52, p<0.05, mean bias±95% CI			**VO_{2max} (L/min)** = 0.98±0.22; **RQ** = 0.95±0.06; **%pHR_{max}** = 84.4±8.4; **Dyspnea mBorg** = 4 (0-10); **Leg effort mBorg** = 4.5 (0-10).
(Simone Dal Corso et al., 2013)	**Criterion Validity** VO_{2peak} in the best step test vs VO_{2peak} in CPET r = 0.79, p < 0.0001	Step test 1 vs. Step test 2 ICC = number of steps 0.99 (0.97-0.99), VO_2 0.99 (0.98-0.995)		IST-1 **VO_{2max}** - 1,19±0,39L/min HR - 134±21 (88±13% of HRmax) **Dyspneia** - 5,4±2,3 **Fatigue** - 5,5±2,6 IST-2 **VO_{2max}** - 1,20±0,40L/min HR - 132±21 (86±12% of HRmax) **Dyspneia** mBorg - 5,1±2,2 **Fatigue** mBorg - 5,1±2,7
(Swinburn et al., 1985)	**Criterion validity** VO_{2max}PST vs. VO_{2max} CPET R^2 = 0.74; p<0.001.			**VO_{2max} = 1010±190 (mL/min)**
(Teren et al., 2016)	**Criterion validity** step test VO_{2max} vs. CPET VO_{2peak}. r = 0.64; p<0.001			HR - 120 (103-138) bpm **Borg** - 14 (12-150) **VO_{2max}** - 26,4 (25,0-28,1) mL/min/kg
(C. de F. Travensolo et al., 2020)	**Construct Validity** number of steps of 6MST vs. distance walked on 6MWT r = 0.7, p<0.001	Step test 1 vs. Step test 2 ICC= 0.967; 95% CI: 0.766, 0.989; p<0.05		HR_{paak} - 117.3±16.3 bpm **Dyspnea mBorg** - 4.6±2.1

(Vilarinho, Mendes, et al., 2021)	Construct Validity number of steps vs. total distance $0.38 \leq r \leq 0.55$; $p < 0.05$	test 1 vs. test 2 of aCST ICC = 0.48 (0.03 to 0.76) p<0.001		%pHR_{max} = 94.0 ± 6.5%; Blood lactate = 11.3 ± 4.5 mmol/dl; RPE Borg = 18.4 ± 1.5.
(Vieira et al., 2020)	Criterion validity step test VO_{2max} vs. CPET VO_{2peak}. r=0.84, $p < 0.001$			VO_{2peak} - 1018±283 (mL/min) RER_{peak} - 1,11±0,11 %HR_{max}- 84±7% mBorg dyspnea - 5±3 mBorg leg effort - 6±2

RP – rehabilitation program; VO_{2max} – maximal oxygen uptake; HR_{max}– maximal heart rate; %HR_{max} – percentage of maximal heart rate; RER - respiratory exchange ratio; RPE – Rate of Perceived Exertion; Borg – Borg Exertion Scale; mBorg – Modified Borg Exertion Scale; VAS – Visual Analog Scale; %VO_{2max} – percentage of VO_{2max}; VO_{2peak} – peak of maximal oxygen uptake; OSA – Obstructive Sleep Apnea; mL – millilitres; kg – kilogram; min – minute; bpm – beats per minute; ND – Not described;

Appendix 4 – Methological quality assessment for measurement properties.

Reference		Content Validity					Structural Validity	Internal Consistency	Cross cultural validity	Reliability	Measurement error	Criterion validity	Construct Validity		Responsiveness		
		Asking patients			Asking experts												
		Relevance	Compreensiveness	Comprehensability	Relevance	Comprehensiveness							Convergent validity	Known groups validity	Comparison with gold standard	Comparison with other instruments	Comparison before and after intervention
2MST	(Ricci et al., 2019)	D	A	A	I	I	V	D	A	D	V	V	N	N	V	N	N
	(Braghieri et al., 2021)	I	I	I	I	I	A	V	V	V	D	N	V	N	N	A	N
	(De Blok et al., 2006)	I	I	I	I	I	D	D	A	N	N	N	N	N	N	N	D
3MST	(Beaumont et al., 2019)	A	A	A	D	D	V	V	V	V	V	N	V	N	N	V	V
	(Iturian Barrón et al., 2021)	N	N	N	D	D	V	V	V	V	V	N	V	N	N	V	N
	(Hong et al., 2019)	D	D	D	I	I	V	A	A	N	D	V	V	N	V	V	N
	(Murphy et al., 2005)	I	I	I	I	I	D	D	A	N	N	N	N	N	N	N	D
	(Borel et al., 2016)	I	I	I	I	I	D	A	A	N	N	N	N	N	A	N	A
3MSTx1	(Cofre-Bolados et al., 2018)	D	A	A	I	I	V	V	A	N	D	V	V	N	V	V	N

6MST	(S. Dal Corso et al., 2007)	I	D	D	D	D	I	D	N	D	V	N	V	N	D	A	V
	(Magalhães et al., 2020)	I	I	I	D	D	D	V	V	V	V	N	V	N	N	V	V
	(Carvalho et al., 2015)	I	I	I	I	I	V	V	V	V	V	V	N	N	V	A	N
	(Ritt et al., 2021)	I	I	I	I	I	A	A	A	N	V	V	N	N	V	N	N
	(Di Thommazo -Luporini et al., 2015)	I	I	I	I	I	V	A	A	N	V	V	N	N	V	N	N
	(Reychler et al., 2018)	I	I	I	I	I	V	A	A	N	V	N	V	N	N	V	N
	(Arcuri et al., 2016)	I	I	I	I	I	A	A	V	V	A	N	V	D	N	V	N
	(C. de F. Travensolo et al., 2020)	I	I	I	I	I	V	A	V	V	A	N	V	N	N	V	N
	(Munari et al., 2020)	D	D	D	I	I	A	A	A	V	V	N	N	N	N	N	N
	(da Costa et al., 2014)	D	D	D	D	D	A	V	V	V	V	N	N	N	N	N	N
	(Pessoa et al., 2014)	D	D	D	I	I	A	A	A	N	N	N	V	N	N	N	N
	(Marrara et al., 2012)	D	D	D	D	D	A	A	A	N	N	N	N	N	D	D	D
6MSrT	(Grosbois et al., 2016)	I	I	I	I	I	A	A	A	N	A	V	V	V	V	V	V
	(Bonnevie et al., 2019)	I	I	I	I	I	V	A	A	N	N	A	A	A	N	N	N

	(Fabre et al., 2017)	I	I	I	I	I	V	A	A	N	N	A	A	A	N	N	N
	(Bonnevie et al., 2017)	I	I	I	I	I	A	V	A	N	N	A	N	N	N	N	N
	(Pichon et al., 2016)	I	I	I	I	I	V	A	A	N	V	N	V	V	V	V	V
	(Coquart et al., 2016)	I	I	I	I	I	A	A	A	N	N	N	N	N	A	A	A
	(Coquart et al., 2015)	I	I	I	I	I	V	V	V	V	A	N	N	N	V	N	V
aCST	(Vilarinho, Mendes, et al., 2021)	I	I	I	D	D	A	A	N	A	D	N	V	N	N	V	N
CST	(Karloh et al., 2013)	N	N	N	I	I	A	A	V	N	V	N	V	V	N	V	N
	(de Camargo et al., 2011)	D	D	D	A	A	A	V	V	V	V	A	V	N	A	V	V
	(J. Buckley et al., 2004)	I	I	I	I	I	V	A	A	V	V	V	N	N	V	N	N
	(José & Corso, 2016)	I	I	I	I	I	V	V	A	N	V	N	V	N	N	V	A
	(Reed et al., 2020)	I	I	I	D	D	A	A	V	N	V	V	N	N	V	A	V
FRST	(Hansen et al., 2011)	I	I	I	D	D	V	V	A	N	I	V	N	N	V	N	A
IST	(Simone Dal Corso et al., 2013)	D	D	D	D	D	V	A	A	V	D	N	V	N	N	V	V
	(Vieira et al., 2020)	I	I	I	I	I	V	V	V	N	V	V	N	N	V	N	N
MIST	(José & Corso, 2016)	I	I	I	I	I	V	V	A	N	V	N	V	N	N	V	A

PST	(Swinburn et al., 1985)	I	I	I	I	I	A	A	A	N	D	V	N	N	N	N	N
SST	(Moore et al., 2013)	N	N	N	D	I	A	D	V	V	V	V	V	V	I	V	V
	(Lemanska et al., 2019)	I	I	I	I	I	V	V	V	V	V	V	N	V	V	N	N
STEP	(Knight et al., 2014)	I	I	I	I	I	A	A	V	N	D	V	N	N	V	N	N
YMCA	(Teren et al., 2016)	I	I	I	I	I	V	D	V	N	V	V	N	N	V	N	N
YMCA-+imHBC	(Beutner et al., 2015b)	I	I	I	I	I	V	A	A	N	D	A	N	D	V	N	N

2MST – 2-Minute Step Test; 3MST – 3-Minute Step Test; 6MST – 6-min Step Test; 6MSrT – 6-minute Stepper Test; 12MWT – 12 Minute Walking Test; aCST – Adapted Chester Step Test; CST – Chester Step Test; IST – Incremental Step Test; mIST – Modified Incremental Step Test; PST – Fixed Rate Paced Step Test; SST – Siconolfi Step Test; SFRST – Standardized Fixed-Rate Step Test; ST3x1 – Step Test 3x1; STEP – Step Test and Exercise Prescription (STEP); YMCA – YMCA-Step Test; +1mHBC – +1-Minute Heart Beat Count.

Appendix 5 – Quality level assessment (GRADE) of the measurement properties.

			Content validity	Relevance	Comprehe nsi-veness	Comprehensibility	Structural validity	Internal consistency	Cross-cultural validity	Measurement invariance	Reliability	Measurement error	Criterion validity	Construct validity	Responsiveness
2MST	OVERALL RATING	+/-/?	-	-	-	-	+	-	+	-	-	-	-	-	-
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	High	Low	High	Low	Low	Low	Low	Low	Very Low
3MST	OVERALL RATING	+/-/?	-	-	-	-	+	+	+	+	+	+	-	+	+
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	High	High	High	Moderate	Moderate	Moderate	Low	High	Moderate
3MST x1	OVERALL RATING	+/-/?	-	-	-	-	+	+	+	?	-	?	+	+	?
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	High	High	High	Very Low	Low	Very Low	High	High	Very Low
6MST	OVERALL RATING	+/-/?	-	-	-	-	+	+	+	?	+	+	-	+	-
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	High	Moderate	High	Low	Moderate	Moderate	Low	Moderate	Very Low
6MSrT	OVERALL RATING	+/-/?	-	-	-	-	+	+	+	?	+	+	-	+	-
	QUALITY OF	High, moderate, low,	Low	Low	Low	Low	High	Moderate	High	Low	Moderate	Moderate	Low	Moderate	Low

	EVIDENCE	very low													
aCST	OVERALL RATING	+/–/?	–	–	–	–	+	+	+	?	–	+	+	–	–
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	High	High	Moderate	Very Low	Low	Moderate	Moderate	Low	Low
CST	OVERALL RATING	+/–/?	–	–	–	–	+	+	+	+	+	+	+	+	–
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	Moderate	Moderate	High	Moderate	Moderate	High	Moderate	Moderate	Low
SFRST	OVERALL RATING	+/–/?	–	–	–	–	+	+	?	?	?	–	+	?	?
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	High	High	Very Low	Very Low	Very Low	Low	High	Very Low	Very Low
IST	OVERALL RATING	+/–/?	–	–	–	–	+	+	?	?	+	?	?	+	+
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	High	High	Very Low	Very Low	High	Very Low	Very Low	High	High
mIST	OVERALL RATING	+/–/?	–	–	–	–	+	+	+	+	–	+	–	+	–
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	High	High	High	High	Low	High	Low	High	Low
PST	OVERALL RATING	+/–/?	–	–	–	–	–	–	–	?	?	?	+	?	?

	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	Low	Low	Low	Low	Very Low	Very Low	Moderate	Very Low	Very Low
SST	OVERALL RATING	+/–/?	–	–	–	–	+	+	–	–	+	+	?	–	–
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	High	High	Moderate	Moderate	High	High	Moderate	Moderate	Low
STEP	OVERALL RATING	+/–/?	–	–	–	–	+	+	+	?	–	?	+	–	–
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	High	High	High	Very Low	Low	Very Low	High	Low	Low
YMCA	OVERALL RATING	+/–/?	–	–	–	–	+	–	+	–	–	–	+	–	–
	QUALITY OF EVIDENCE	High, moderate, low, very low	Low	Low	Low	Low	High	Low	High	Low	Low	Low	High	Low	Low

2MST – 2-Minute Step Test; 3MST – 3-Minute Step Test; 6MST – 6-min Step Test; 6MSrT – 6-minute Stepper Test; 12MWT – 12 Minute Walking Test; aCST – Adapted Chester Step Test; CST – Chester Step Test; IST – Incremental Step Test; mIST – Modified Incremental Step Test; PST – Fixed Rate Paced Step Test; SST – Siconolfi Step Test; SFRST – Standardized Fixed-Rate Step Test; ST3x1 – Step Test 3x1; STEP – Step Test and Exercise Prescription (STEP); YMCA – YMCA-Step Test; +1mHBC – +1-Minute Heart Beat Count.

Appendix 6 – Level of quality of the study designs.

Article	**A: Selection Bias**	**B: Study Design**	**C: Confounders**	**D: Blinding**	**E: Data Collection Methods**	**F: Withdrawals and Drop-outs**	**Level of quality**
(Beaumont et al., 2019)	Strong	Moderate	Weak	Strong	Strong	Strong	**Moderate**
(Beutner et al., 2015b)	Strong	Weak	Moderate	Weak	Moderate	Strong	**Weak**
(Bonnevie et al., 2017)	Moderate	Weak	Strong	Weak	Strong	Strong	**Weak**
(Borel et al., 2016)	Moderate	Weak	Moderate	Weak	Strong	Strong	**Weak**
(C. de F. Travensolo et al., 2020)	Strong	Strong	Strong	Weak	Strong	Strong	**Moderate**
(Carvalho et al., 2015)	Moderate	Weak	Strong	Weak	Strong	Strong	**Weak**
(Coquart et al., 2015)	Moderate	Moderate	Strong	Weak	Strong	Strong	**Moderate**
(Coquart et al., 2016)	Weak	Strong	Moderate	Strong	Strong	Strong	**Moderate**
(de Camargo et al., 2011)	Moderate	Strong	Weak	Strong	Strong	Strong	**Weak**
(Di Thommazo-Luporini et al., 2015)	Moderate	Weak	Strong	Weak	Moderate	Strong	**Weak**
(Fabre et al., 2017)	Moderate	Weak	Strong	Weak	Moderate	Moderate	**Weak**
(Hansen et al., 2011)	Strong	Weak	Weak	Weak	Moderate	Weak	**Weak**
(Hong et al., 2019)	Strong	Moderate	Strong	Weak	Strong	Strong	**Moderate**
(Iturian Barrón et al., 2021)	Weak	Strong	Moderate	Strong	Strong	Strong	**Moderate**
(J. Buckley et al., 2004)	Weak	Weak	Weak	Weak	Strong	Strong	**Weak**
(José & Corso, 2016)	Moderate	Moderate	Strong	Weak	Strong	Strong	**Moderate**
(Knight et al., 2014)	Strong	Moderate	Strong	Weak	Moderate	Strong	**Moderate**
(Lemanska et al., 2019)	Moderate	Weak	Strong	Weak	Strong	Moderate	**Weak**

(Karloh et al., 2013)	Weak	Strong	Weak	Weak	Moderate	Strong	**Weak**
(Magalhães et al., 2020)	Strong	Strong	Strong	Weak	Strong	Moderate	**Moderate**
(Marrara et al., 2012)	Moderate	Weak	Weak	Moderate	Moderate	Moderate	**Weak**
(Moore et al., 2013)	Strong	Weak	Weak	Weak	Strong	Moderate	**Weak**
(Munari et al., 2020)	Moderate	Weak	Weak	Weak	Strong	Weak	**Weak**
(Murphy et al., 2005)	Moderate	Weak	Moderate	Weak	Moderate	Moderate	**Weak**
(Pichon et al., 2016)	Moderate	Weak	Weak	Weak	Strong	Weak	**Weak**
(Reychler et al., 2018)	Strong	Strong	Strong	Moderate	Moderate	Strong	**Moderate**
(Ricci et al., 2019)	Moderate	Weak	Weak	Moderate	Moderate	Moderate	**Weak**
(Ritt et al., 2021)	Moderate	Weak	Moderate	Weak	Moderate	Moderate	**Weak**
(S. Dal Corso et al., 2007)	Moderate	Weak	Weak	Weak	Strong	Moderate	**Weak**
(Simone Dal Corso et al., 2013)	Moderate	Weak	Weak	Weak	Strong	Weak	**Weak**
(Teren et al., 2016)	Strong	Weak	Weak	Weak	Moderate	Strong	**Weak**
(Vieira et al., 2020)	Moderate	Moderate	Weak	Weak	Moderate	Strong	**Weak**
(Vilarinho, Mendes, et al., 2021)	Weak	Weak	Weak	Weak	Moderate	Moderate	**Weak**