Validity and Reproducibility of a New Treadmill Protocol: The Fitkids Treadmill Test

ELLES M. W. KOTTE[1], JANKE F. DE GROOT[2,3,4], BART C. BONGERS[2,3,4,5], ALEXANDER M. F. WINKLER[1], and TIM TAKKEN[3,4]

[1]Fitkids Foundation, Amsterdam, THE NETHERLANDS; [2]University of Applied Sciences, Utrecht, THE NETHERLANDS; [3]Child Development and Exercise Center, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, THE NETHERLANDS; [4]Partner of Shared Utrecht Pediatric Exercise Research Laboratory, Utrecht, THE NETHERLANDS; and [5]Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, THE NETHERLANDS

ABSTRACT

KOTTE, E. M. W., J. F. DE GROOT, B. C. BONGERS, A. M. F. WINKLER, and T. TAKKEN.. Validity and Reproducibility of a New Treadmill Protocol: The Fitkids Treadmill Test. *Med. Sci. Sports Exerc.*, Vol. 47, No. 10, pp. 2241–2247, 2015. **Purpose**: This study aimed to investigate the validity and reproducibility of a new treadmill protocol in healthy children and adolescents: the Fitkids Treadmill Test (FTT). **Methods**: Sixty-eight healthy children and adolescents (6–18 yr) were randomly divided into a validity group (14 boys and 20 girls; mean ± SD age, 12.9 ± 3.6 yr) that performed the FTT and Bruce protocol, both with respiratory gas analysis within 2 wk, and a reproducibility group (19 boys and 15 girls; mean ± SD age, 13.5 ± 3.5 yr) that performed the FTT twice within 2 wk. A subgroup of 21 participants within the reproducibility group performed both FTT with respiratory gas analysis. Time to exhaustion (TTE) was the main outcome of the FTT. **Results**: $\dot{V}O_{2peak}$ measured during the FTT showed excellent correlation with $\dot{V}O_{2peak}$ measured during the Bruce protocol ($r = 0.90$; $P < 0.01$). Backward multiple regression analysis provided the following prediction equations for $\dot{V}O_{2peak}$ (L·min^{-1}) for boys and girls, respectively: $\dot{V}O_{2peak}$ FTT $= -0.748 + (0.117 \times TTE FTT) + (0.032 \times body\,mass) + 0.263$, and $\dot{V}O_{2peak}$ FTT $= -0.748 + (0.117 \times TTE FTT) + (0.032 \times body\,mass)$ [$R^2 = 0.935$, SEE $= 0.256\,L·min^{-1}$]. Cross-validation of the regression model showed an R^2 value of 0.76. Reliability statistics for the FTT showed an intraclass correlation coefficient of 0.985 (95% confidence interval, 0.971–0.993; $P < 0.001$) for TTE. Bland–Altman analysis showed a mean bias of -0.07 min, with limits of agreement between $+1.30$ and -1.43 min. **Conclusions**: Results suggest that the FTT is a useful treadmill protocol with good validity and reproducibility in healthy children and adolescents. Exercise performance on the FTT and body mass can be used to adequately predict $\dot{V}O_{2peak}$ when respiratory gas analysis is not available. **Key Words**: EXERCISE TESTING, PHYSICAL FITNESS, RELIABILITY, CHILD

Standardized exercise testing remains an important tool that provides valuable diagnostic and prognostic information in daily clinical practice. Exercise testing allows individualized assessment of exercise tolerance and can be used to monitor exercise training programs. For use in daily clinical practice, nonsophisticated, inexpensive, reliable, and valid exercise tests are of increasing interest, as they might help increase the use of exercise testing (6).

The Bruce treadmill protocol is the most frequently used protocol in children and adolescents using a treadmill for cardiopulmonary exercise testing (CPET) (12), and pediatric norm values have been published (4,30,31). Although the Bruce treadmill protocol has good validity and reproducibility, the use in outpatient physical therapy practices can be difficult because the test requires a treadmill ergometer that can operate at an incline of 22%. Many of these practices are embedded in health and sports centers with only standard treadmills available that can operate at a maximum incline of 15%. Therefore, the angle of inclination of the Bruce treadmill protocol is a major concern in outpatient physical therapy practices.

Although there are several other established maximal treadmill protocols, including the Balke protocol (3), the Cornell protocol (21), and the German Society of Pediatric Cardiology protocol (18), the maximum incline of these treadmill protocols exceeds 15% as well. In addition to this practical issue, many of the established maximal treadmill protocols are too demanding for children with disability or chronic disease because of the high incline in the first stage of the protocol, leading to premature exhaustion of the muscles of the lower limbs before achieving cardiac or respiratory limits (23).

In summary, there is a need for a maximal treadmill protocol that can be used in outpatient physical therapy practices when limited to a treadmill machine with maximal incline of 15%. In this article, a new maximal treadmill protocol, the so-called Fitkids Treadmill Test (FTT), is introduced. The protocol starts with 0° incline, making this protocol useful in children and adolescents with disability or chronic disease. Before application in clinical practice, insight in the clinimetric properties of the FTT is crucial. The current study is the first step in the identification of the clinimetric properties of the FTT and aimed to investigate the validity and reproducibility of the new developed FTT in healthy children and adolescents between 6 and 18 yr of age.

METHODS

Participants

Healthy children and adolescents (6–18 yr of age) were recruited from primary and secondary schools as well as from different sports clubs located in the Netherlands (convenience sample). The inclusion of participants started after approval of the Central Committee on Research Involving Human Subjects in the Netherlands. The participants and their parents individually received written information about the study and informed consent forms. Signed informed consent forms were obtained from the parents or legal guardian of each participant and separately from each participant who was 12 yr or older. The modified physical activity readiness questionnaire (PAR-Q) was used to evaluate the health status of willing participants and to assess safety for performing maximal exercise. Exclusion criteria were as follows: use of medication affecting exercise capacity, cardiovascular, or respiratory disease, impaired motor development, morbid obesity (body mass index (BMI), >2.5 SDS), or positive responses to one or more of the modified PAR-Q questions.

Study Design

Participants were randomly divided into a validity or a reproducibility group. To assess the validity of the FTT, the validity group performed the FTT and the Bruce test with respiratory gas analysis in a counterbalanced order within 2 wk. To assess the reproducibility of the FTT, the reproducibility group performed the FTT twice within 2 wk. A subgroup of 21 participants within the reproducibility group performed both FTT with respiratory gas analysis.

Anthropometry

Body mass and height were determined to the nearest 0.5 kg and 0.1 cm, respectively, using an analog scale (Medisana PSD; Medisana Benelux N.V.; Kerkrade, the Netherlands) and a stadiometer (Seca 213; Seca, Hamburg, Germany), respectively. For both measurements, participants wore light clothes and no shoes. BMI was derived from body mass and height. The BMI for age SD scores was calculated using Dutch reference values (28). Subcutaneous fat of the biceps, triceps, subscapular, and suprailiac regions was measured using a Harpenden skinfold caliper. The average of three measures of each area was used. Body density was estimated using the equations proposed by Deurenberg et al. (17) and used to estimate percent body fat on the basis of the Siri equation (32). Body surface area (BSA) was calculated using the equation of Haycock et al. (20), which has been validated in infants, children, and adults.

Exercise Testing

Exercise tests were performed on a motor-driven treadmill ergometer (Lode Valiant; Lode BV, Groningen, the Netherlands) using the Lode Ergometry Manager software (Lode BV, Groningen, the Netherlands). During all treadmill tests, heart rate (HR) was monitored using a soft strap with an HR sensor (Polar H1 transmitter; Polar, Kempele, Finland). The participants of the validity group and the 21 participants of the reproducibility subgroup breathed through a face mask (Hans Rudolph Inc., Kansas City, MO) during both maximal treadmill tests, which was connected to a mobile gas analysis system (Cortex Metamax B^3; Cortex Medical GmbH, Leipzig, Germany) with a built-in gas analyzer. The mobile respiratory gas analysis system was calibrated for respiratory gas analysis measurements (ambient air and a gas mixture of 17% oxygen and 5% carbon dioxide) and volume measurements (3-L syringe). Values for oxygen uptake ($\dot{V}O_2$), carbon dioxide production ($\dot{V}CO_2$), minute ventilation ($\dot{V}E$), and RER were collected at 10-s intervals. The Cortex Metamax B^3 is a valid and reliable system for measuring ventilatory parameters during exercise (9,25,26).

Peak $\dot{V}O_2$ ($\dot{V}O_{2peak}$), peak $\dot{V}E$ ($\dot{V}E_{peak}$), and peak respiratory exchange ratio (RER) (RER_{peak}), measured as the average value over the last 30 s before peak exercise, and peak HR (HR_{peak}), defined as the highest value achieved during the last 30 s before test termination, were used for analysis. The test was deemed maximal when at least one of the following criteria was met: an HR_{peak} >180 bpm or an RER_{peak} >1.0 (1). Before and directly after the exercise test, participants were asked to rate their perceived exhaustion on a 10-point visual analog scale (VAS). Perceived exertion of the FTT and the Bruce test were determined by subtracting the pretest VAS score from the posttest VAS score (ΔVAS; posttest VAS score minus pretest VAS score).

FTT

The FTT protocol consists of a 90-s warm-up period (3.5 km·h^{-1}, 0% grade) followed by the initiation of the test at 3.5 km·h^{-1} and a 1% gradient for 90 s followed by incremental increases in both speed (0.5 km·h^{-1}) and incline (2%) every 90 s until an incline of 15% was attained (Table 1). After this last step, the incline is held at 15% and incremental increases of speed (0.5 km·h^{-1}) are performed every 90 s until

volitional exhaustion. After the test, participants are monitored for 2 min to ensure normal recovery of HR (2.0 km·h^{-1} with a flat treadmill). The main outcome measure of the FTT is time to exhaustion (TTE) and is defined as the point at which the participant chooses to stop despite strong verbal encouragement. TTE is calculated as the total duration of the test minus the duration of the warm-up phase.

Bruce Test

The Bruce treadmill protocol consisted of a 90-s warm-up period (2.74 km·h^{-1} and a flat treadmill) followed by the initiation of the test at 2.74 km·h^{-1} and a 10% gradient for 3 min followed by incremental increases in speed and incline every 3 min until volitional exhaustion, as described elsewhere (10). After the test, participants were monitored for 2 min to ensure a normal recovery of HR (2.0 km·h^{-1} with a flat treadmill). The main outcome measure of the Bruce protocol is TTE and is defined as the point at which the participant chooses to stop despite strong verbal encouragement. TTE is calculated as the total duration of the test minus the duration of the warm-up phase.

During both tests, participants are not allowed to use the handrails, except for touching the handrail with one or two fingers to maintain body position near the center of the moving belt.

Statistical Analysis

The IBM SPSS Statistics for Windows version 20.0 (IBM Corp., Armonk, NY) was used for data analyses. The distribution of the variables was assessed using visual inspection (histogram, boxplot, and normal Q–Q plot) and the Shapiro–Wilk test for normality. Except for ΔVAS scores in both groups and $\dot{V}E_{peak}$ in the validity group, all variables were normally distributed, with skewness and kurtosis z-scores between −1.96 and +1.96.

For both the validity group and the reproducibility group, data were checked on significant differences for TTE, cardiopulmonary variables, and perceived exertion between both tests. Paired-samples t-tests were completed for normally distributed data, whereas Wilcoxon signed-rank tests were completed for nonnormally distributed data.

TABLE 1. Protocol FTT.

Stage	Stage Duration (s)	Speed (km·h^{-1})	Speed (mph)	Gradient (%)
Warm-up	90	3.5	2.175	0
1	90	3.5	2.175	1
2	90	4.0	2.485	3
3	90	4.5	2.796	5
4	90	5.0	3.106	7
5	90	5.5	3.417	9
6	90	6.0	3.728	11
7	90	6.5	4.039	13
8	90	7.0	4.350	15
9	90	7.5	4.661	15
…				

Speed is not limited until voluntary exhaustion.
Maximum gradient is limited to 15%.

Validity. To examine the validity of the FTT, correlation coefficients were calculated between TTE achieved at the FTT and TTE achieved at the Bruce test as well as between the cardiopulmonary variables achieved at both tests. Backward multiple linear regression analysis was used to generate an equation to predict $\dot{V}O_{2peak}$ achieved at the FTT. Pearson product correlations were calculated between anthropometric variables (age, height, body mass, percent body fat, fat-free mass (FFM), BMI, and BSA) and $\dot{V}O_{2peak}$ attained at the FTT. The best predictor candidate for $\dot{V}O_{2peak}$ attained at the FTT was entered into the model along with sex (0 = girls, 1 = boys) and TTE of the FTT. The goodness of fit and precision of the regression equation were evaluated using multiple coefficient of determination (R^2) and the absolute SEE. The model was tested for random errors, homoscedasticity, and multicollinearity. The regression model was cross-validated using the data from the respiratory gas analysis obtained in the reproducibility subgroup ($n = 21$). Before the cross-validation analysis, the validity group and the reproducibility group were checked on significant differences in participant characteristics. Independent-sample t-tests were completed for normally distributed data, and Mann–Whitney Tests were performed for nonnormally distributed data.

Reproducibility. For test–retest reliability analysis, the two-way mixed (type, absolute agreement) intraclass correlation coefficient (ICC) was calculated for TTE achieved at both FTT and for cardiopulmonary variables achieved at both FTT. ICC values higher than 0.75 were considered acceptable (27). To determine whether improvements are meaningful for a single patient, limits of agreement (LOA) were calculated for the main outcome measure of the FTT, TTE, according to the procedure described by Bland and Altman (5). TTE was defined as the main outcome measure in the reproducibility analysis because many clinicians might not be able to collect expired gases during exercise. A P value < 0.05 was considered statistically significant.

RESULTS

Of the 74 participants approached for this study, none were excluded because of a positive response in the PAR-Q. The study population was randomly divided into a validity group and a reproducibility group. From both the validity group and the reproducibility group, three participants were excluded because of the following reasons: invalid respiratory gas analysis ($n = 1$), illness during the second day of testing ($n = 2$), dyspnea during the test ($n = 1$), scheduling issues ($n = 1$), and a painful leg during testing ($n = 1$). Eventually, the validity group and reproducibility group both consisted of 34 participants. Participant descriptive statistics are found in Table 2.

Validity. Participants from the validity group performed both the FTT and the Bruce protocol without any adverse effects. They all met the subjective criteria of maximal effort (sweating, unsteady walking, facial flushing, and clear unwillingness to

TABLE 2. Participant characteristics.

	Validity Group	Reproducibility Group	P Value
Sex (boys/girls)	14/20	19/15	
Age (yr)	12.9 ± 3.6 (6.5 to 18.6)	13.5 ± 3.5 (6.6 to 18.2)	0.501
Body mass (kg)	48.9 ± 16.8 (20.0 to 80.8)	49.2 ± 15.9 (21.0 to 80.5)	0.928
Body height (m)	1.57 ± 0.18 (1.15 to 1.85)	1.61 ± 0.19 (1.20 to 1.89)	0.374
BMI (kg·m^{-2})	19.1 ± 3.2 (13.6 to 25.3)	18.3 ± 2.3 (12.9 to 22.5)	0.258
BSA (m^2)	1.44 ± 0.34 (0.80 to 2.04)	1.47 ± 0.33 (0.83 to 2.05)	0.775
FFM (kg)	39.1 ± 13.0 (17.2 to 64.0)	40.1 ± 13.2 (16.7 to 61.9)	0.750
Body fat (%)	19.4 ± 4.5 (12.8 to 27.2)	18.6 ± 4.2 (10.8 to 26.1)	0.486
BMI for age (SD score)	−0.02 (−0.45 to 0.37)	0.30 (−0.20 to 1.24)	0.048*[a]

Date are presented as mean ± SD (range).
[a] Nonparametric Mann–Whitney Test, data presented as median (interquartile range).
*$P < 0.05$.

continue despite strong verbal encouragement) and the objective criteria of maximal effort ($HR_{peak} > 180$ bpm and/or $RER_{peak} > 1.0$) during both the FTT and the Bruce treadmill protocol. Seventeen of the 34 participants (50%) reached the maximal incline of 15%. The mean ± SD between-visit time is 6.2 ± 1.4 d.

Results of the FTT and the Bruce protocol completed by the validity group are presented in Table 3. Significant higher values of TTE were found for the FTT. The cardiopulmonary variables and rate of perceived exertion were not significantly different between the FTT and the Bruce protocol. Pearson correlation coefficients for TTE, HR_{peak}, $\dot{V}O_{2peak}$ (L·min^{-1}), and RER_{peak} achieved at the Bruce protocol and the FTT were 0.98, 0.67, 0.90, and 0.64 (all $P < 0.01$), respectively, which are categorized as moderate-to-strong correlations according to Dancey and Reidy's categorization (14). For $\dot{V}E_{peak}$, a Spearman correlation coefficient of 0.96 was found ($P < 0.01$).

To construct an equation to predict $\dot{V}O_{2peak}$ attained at the FTT, Pearson correlation coefficients were calculated between anthropometric variables and $\dot{V}O_{2peak}$ attained at the FTT. FFM seemed to be the best predictor candidate to include into the regression model ($r = 0.934$, $P = 0.01$). However, FFM is impractical for daily use and was therefore not included in the final model. Body mass, which was the second best predictor and easily available in daily practice, was included in the final model along with sex and TTE on the FTT. The model that incorporated FFM can be found in the supplemental digital content (see Document, Supplemental Digital Content 1, Equations to predict $\dot{V}O_{2peak}$ achieved during the FTT from the attained TTE at the FTT and FFM, http://links.lww.com/MSS/A530).

The following equations for boys and girls were generated to predict $\dot{V}O_{2peak}$ achieved at the FTT from the attained TTE on the FTT and body mass:

$$\text{Boys}: \dot{V}O_{2peak}\,\text{FTT} = -0.748 + (0.117 \times \text{TTE FTT}) \\ + (0.032 \times \text{body mass}) + 0.263$$

$$\text{Girls}: \dot{V}O_{2peak}\,\text{FTT} = -0.748 + (0.117 \times \text{TTE FTT}) \\ + (0.032 \times \text{body mass})$$

In the prediction equations, "$\dot{V}O_{2peak}$ FTT" represents the predicted $\dot{V}O_{2peak}$ in liters per minute, "TTE FTT" is the total duration of the FTT in minutes minus the duration of the warm-up phase in minutes, and "body mass" is expressed in kilograms ($R^2 = 0.935$, SEE = 0.256).

For cross-validation purposes, the $\dot{V}O_{2peak}$ estimated by the equations was plotted against the measured $\dot{V}O_{2peak}$ during the first FTT in the reproducibility subgroup who performed the FTT with respiratory gas analysis (Fig. 1). Estimated $\dot{V}O_{2peak}$ (3.19 ± 0.51 L·min^{-1}) did not differ significantly from the observed $\dot{V}O_{2peak}$ (3.31 ± 0.88 L·min^{-1}; $P = 0.291$; $R^2 = 0.76$).

Reproducibility. Participants from the reproducibility group performed the FTT twice without any adverse effects. They all met the subjective criteria of maximal effort (sweating, unsteady walking, facial flushing, and clear unwillingness to continue despite strong verbal encouragement) and the objective criteria of maximal effort ($HR_{peak} > 180$ bpm and/or $RER_{peak} > 1.0$) during both FTT. One participant showed HR_{peak} of 174 and 169 bpm, respectively, on the first and second FTT; however, RER_{peak}

TABLE 3. FTT and Bruce protocol results of the validity group.

	FTT ($n = 34$)	Bruce Protocol ($n = 34$)	P Value
TTE (min)	13.5 ± 3.7 (9.2 to 22.2)	12.1 ± 3.0 (7.9 to 18.0)	<0.001*
HR_{peak} (bpm)	194.9 ± 6.9 (181 to 208)	194.1 ± 6.3 (183 to 207)	0.360
$\dot{V}O_{2peak}$ (L·min^{-1})	2.5 ± 1.0 (1.1 to 4.4)	2.5 ± 1.0 (1.0 to 4.5)	0.779
$\dot{V}O_{2peak}$ (mL·kg^{-1}·min^{-1})	51.1 ± 7.6 (28.0 to 71.0)	51.8 ± 8.2 (34.0 to 72.0)	0.366
RER_{peak}	1.1 ± 0.1 (1.0 to 1.3)	1.1 ± 0.1 (1.0 to 1.3)	0.947
$\dot{V}E_{peak}$ (L·min^{-1})	75.3 (56.3 to 96.6)	71.09 (55.0 to 103.3)	0.918[a]
ΔVAS	10 (9.0 to 10.0)	9.5 (8.75 to 10)	0.491[a]

Date are presented as mean ± SD (range).
[a] Nonparametric Wilcoxon signed-ranks test; data are presented as median (interquartile range).
*$P \leq 0.001$.
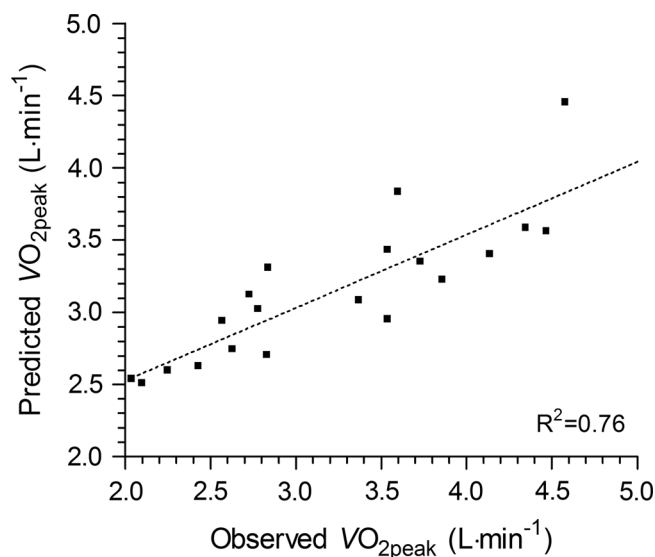ΔVAS, VAS difference addressing the participants' level of fatigue (posttest VAS score minus pretest VAS score).

FIGURE 1—Cross-validation plot: linear relation between the predicted $\dot{V}O_{2peak}$ values based on the prediction equation of the regression model and the $\dot{V}O_{2peak}$ measured during the first FTT in the reproducibility subgroup.

FIGURE 2—Bland–Altman plot of TTE as attained at the first FTT versus the second FTT.

measured during these tests were 1.28 and 1.25, respectively. Twenty-two of the 34 participants (65%) reached the maximal incline of 15%. The mean ± SD between-visit time was 8.9 ± 3.8 d.

The results of both FTT performed by the reproducibility group are shown in Table 4. There were no significant differences in TTE and HR_{peak} between the two FTT. In addition, for perceived exertion (ΔVAS), no significant difference was found between the two FTT. For the cardiopulmonary variables, a significant higher RER_{peak} was found during the second FTT. To quantify the relation between TTE achieved at both tests and between the cardiopulmonary variables achieved at both tests, ICC values were calculated. The ICC for TTE, which is the main outcome for the test–retest reliability statistics, was 0.985 (95% confidence interval (CI), 0.971–0.993; $P < 0.001$). The ICC values for HR_{peak}, $\dot{V}O_{2peak}$, RER_{peak}, and $\dot{V}E_{peak}$ were 0.767 (95% CI, 0.584–0.876; $P < 0.001$), 0.963 (95% CI, 0.912–0.985; $P < 0.001$), 0.631 (95% CI, 0.269–0.834; $P < 0.001$), and 0.948 (95% CI, 0.877–0.978; $P < 0.001$), respectively. To analyze agreement between the two FTT, a Bland–Altman
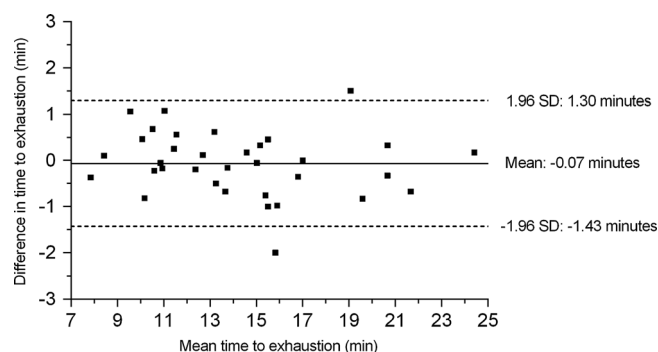
plot for TTE is depicted in Figure 2. The mean bias between the two FTT was −0.07 min. The LOA for TTE were +1.30 and −1.43 min.

DISCUSSION

The aim of the current study was to investigate the validity and reproducibility of a new maximal treadmill protocol for children and adolescents between 6 and 18 yr of age: the FTT. The results of the present study suggest that the FTT is a useful treadmill protocol with good validity and reproducibility. Moderate-to-strong significant correlations between TTE and cardiopulmonary variables attained at the FTT and the Bruce protocol were found, indicating that the FTT is a valid test for the assessment of aerobic exercise capacity in healthy children and adolescents. At the same time, there were no significant differences between the $\dot{V}O_{2peak}$ measured during the Bruce protocol and the $\dot{V}O_{2peak}$ measured during the FTT. These results suggest that the tests produce quite similar estimates and that the tests may be interchangeable because of these similar results.

The current study showed that $\dot{V}O_{2peak}$ can be adequately predicted from TTE and body mass in boys and girls, explaining 94% of the total variance in $\dot{V}O_{2peak}$. When applying the multiple prediction equations developed in the validity group to the reproducibility subgroup, no significant difference was found between the observed $\dot{V}O_{2peak}$ (3.31 ± 0.88 L·min^{-1}) and the $\dot{V}O_{2peak}$ estimated from the prediction equation (3.19 ± 0.51 L·min^{-1}; $P = 0.291$).

TABLE 4. FTT results of the reproducibility group.

		First FTT	Second FTT	P Value
TTE (min)	n = 34	14.3 ± 4.0 (7.7 to 24.5)	14.3 ± 4.1 (8.0 to 24.3)	0.584
HR_{peak} (bpm)	n = 34	197 ± 8 (174 to 215)	196 ± 10 (169 to 222)	0.157
$\dot{V}O_{2peak}$ (L·min^{-1})	n = 21	3.3 ± 0.9 (2.0 to 5.1)	3.3 ± 0.8 (2.2 to 5.0)	0.645
$\dot{V}O_{2peak}$ (mL·kg^{-1}·min^{-1})	n = 21	54.5 ± 12.7 (26.0 to 80.0)	55.2 ± 10.1 (40.0 to 77.0)	0.644
RER_{peak}	n = 21	1.1 ± 0.1 (1.0 to 1.31)	1.2 ± 0.1 (1.0 to 1.4)	0.029*
$\dot{V}E_{peak}$ (L·min^{-1})	n = 21	105.8 ± 25.8 (64.4 to 152.2)	107.3 ± 27.0 (65.7 to 151.2)	0.407
ΔVAS	n = 34	10 (8.0 to 10.0)	10 (8.38 to 10.0)	0.241^a

Date are presented as mean ± SD (range).
^{a}Nonparametric Wilcoxon signed-ranks test; data are presented as median (interquartile range).
*$P < 0.05$.
ΔVAS, VAS difference addressing the participants' level of fatigue (posttest VAS score minus pretest VAS score).

The multiple correlation of the cross-validation regression equation was good ($R^2 = 0.76$). This means that, for clinicians who do not have the resources to directly measure $\dot{V}O_{2peak}$, the TTE on the FTT is an adequate alternative that gives insight in the aerobic fitness of healthy children and adolescents. Nevertheless, measuring $\dot{V}O_{2peak}$ using respiratory gas analysis during incremental exercise is still considered the gold standard for aerobic fitness by the World Health Organization (29).

Several other studies have predicted aerobic fitness from functional performance during exercise testing, both in adults and the pediatric population. Bruce et al. (10) developed the first predictive equations, which were population specific for active and sedentary men with and without cardiac conditions as well as for healthy adults (Pearson correlation coefficients between predicted $\dot{V}O_{2peak}$ and measured $\dot{V}O_{2peak}$ ranged from $r = 0.87$ to $r = 0.92$). Foster et al. (19) later developed a more generalized equation, dependent only on Bruce treadmill test performance, to predict $\dot{V}O_{2peak}$ ($L \cdot min^{-1}$) ($R^2 = 0.98$, SEE = 3.35 $mL \cdot kg^{-1} \cdot min^{-1}$ or 8.5%). This generalized equation was developed for use in patients with cardiac diseases and in healthy sedentary and active individuals. Buono et al. (11) predicted $\dot{V}O_{2peak}$ ($L \cdot min^{-1}$) during a timed distance run on an oval dirt track in a healthy children and adolescents. The equation in their study (based on mile run time, sex, skinfold thickness, and body mass) explained 84% of the total variance in $\dot{V}O_{2peak}$ ($R^2 = 0.84$, SEE = 4.3 $mL \cdot kg^{-1} \cdot min^{-1}$ or 9%). Bongers et al. (7,8) concluded that $\dot{V}O_{2peak}$ ($L \cdot min^{-1}$) could be validly predicted from the attained peak work rate at the steep ramp test in healthy children and adolescents ($R^2 = 0.917$, SEE = 0.24 $L \cdot min^{-1}$ or 9%), and Dencker et al. (16) found that $\dot{V}O_{2peak}$ ($L \cdot min^{-1}$) could be predicted from the peak work rate reached during CPET in healthy children ($R^2 = 0.83$, SEE = 0.11 $L \cdot min^{-1}$ or 8.4%). In the study of Dencker et al. (16), CPET was performed on an electronically braked cycle ergometer using a protocol with an initial workload of 30 W and an increase of 15 $W \cdot min^{-1}$ (1 W every 4 s). The equation developed in the current study is in agreement with previous studies reporting that the addition of body mass (11,24) and sex (11) improves the prediction of $\dot{V}O_{2peak}$ from TTE in children.

Test–retest reproducibility encompasses both reliability and agreement. Whereas the first refers to the consistency or stability of a test after repeated trials, the latter analyzes the variation within the individual scores during a test–retest situation (2,15) and is used to determine the clinical value of a measurement. The test–retest reliability of the FTT can be considered excellent with an ICC for TTE of 0.985 (95% CI, 0.971–0.993; $P < 0.001$). This means that the measurement error is small compared to the variability between the participants and that the discrimination of persons is hardly affected by measurement error. These results are comparable with those reported by Cumming et al. (13). They investigated the test–retest reliability of the Bruce protocol in 20 schoolchildren age 7–13 yr and reported a correlation coefficient of 0.94 between TTE achieved on trial 1 and TTE achieved on trial 2. The mean TTE values in the study of Cumming et al. (13) were 13.9 ± 2.1 min for trial 1 and 13.7 ± 1.9 min for trial 2. The mean TTE values of the FTT observed in the current study were 14.2 ± 4.0 min for trial 1 and 14.3 ± 4.1 min for trial 2. The results of the current study are also comparable with those of Johnston et al. (22) who determined test–retest reproducibility of cardiopulmonary variables during CPET in children using a treadmill protocol. The reported ICC for $\dot{V}O_{2peak}$, HR_{peak}, and $\dot{V}E_{peak}$ were 0.96, 0.87, and 0.91, respectively, against 0.96, 0.77, and 0.95 for $\dot{V}O_{2peak}$, HR_{peak}, and $\dot{V}E_{peak}$, respectively, found in the current study.

The agreement analysis revealed narrow LOA (+1.30 to −1.43 min), indicating that the agreement of the FTT is good. The average difference of TTE attained at the two FTT was roughly 4 s. There was no evidence for a significant learning effect as reflected by a symmetrical distribution of the differences around the zero difference line. A change score in TTE between two concessive measurements within an individual can only be considered to represent a real change if it is outside the LOA.

A limitation of this study is that only healthy participants were tested, so the equations currently developed are appropriate only for healthy children and adolescents. As the FTT is developed for use in outpatient physical therapy practices, predicting models for the clinical population should be developed and evaluated. An additional limitation of the current study is the small sample size used in the cross-validation analysis. The predictive accuracy of the developed regression equation seems proportionally biased, in which true values less than 2.5 $L \cdot min^{-1}$ are systematically overestimated and true values more than 4.0 $L \cdot min^{-1}$ are systematically underestimated. The predictive accuracy of the developed regression equation should be evaluated using a larger sample in future studies. Future studies should also look at reference values for different populations.

In conclusion, the results of the current study suggest that the FTT is a useful treadmill protocol with good validity and reproducibility in healthy children and adolescents age 6–18 yr. The use of the FTT can be favored instead of the Bruce protocol when limited to a standard treadmill machine with maximum incline of 15%. We have shown that exercise performance on the FTT and body mass can be used to simply predict $\dot{V}O_{2peak}$ in healthy children and adolescents in situations where it is not possible to measure $\dot{V}O_{2peak}$ with respiratory gas analysis. Further testing of the FTT in clinical populations is warranted.

REFERENCES

1. Armstrong N, Welsman JR. Aerobic fitness. In: Armstrong N, van Mechelen W, eds. *Paediatric Exercise Science and Medicine*, 2nd ed. Oxford (United Kingdom): Oxford University Press, 2008:97–108.

2. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*. 1998;26(4):217–38.

3. Balke B, Ware RW. An experimental study of physical fitness of Air Force personnel. *U S Armed Forces Med J*. 1959;10(6):675–88.

4. Binkhorst RA, van 't Hof MA, Saris WHM. *Maximale Inspanning Door Kinderen: Referentiewaarden Voor 6–18 Jarige Meisjes en Jongens [Maximal Exercise in Children: Reference Values Girls and Boys, 6 to 18 Year of Age]*. Den Haag (The Netherlands): Brochure Nederlandse Hartstichting; 1992. p. 1–64.

5. Bland JM, Altman DG. Statistical methods for assessing agreement between two models of clinical measurement. *Lancet*. 1986;1:307–10.

6. Bongers BC. *Pediatric Exercise Testing: In Health and Disease*. Maastricht (The Netherlands): Universitaire Pers Maastricht; 2013. pp. 1–182.

7. Bongers BC, de Vries SI, Helders PJ, Takken T. The steep ramp test in healthy children and adolescents: reliability and validity. *Med Sci Sports Exerc*. 2013;45(2):366–71.

8. Bongers BC, Takken T. The paediatric version of the steep ramp test. *J Physiother*. 2014;60:113.

9. Brehm MA, Harlaar J, Groepenhof H. Validation of the portable VmaxST system for oxygen-uptake measurement. *Gait Posture*. 2004;20:67–73.

10. Bruce RA, Kusumi F, Hosmer D. Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease. *Am Heart J*. 1973;85:546–62.

11. Buono JM, Roby JJ, Micale FG, Sallis JF, Shepard E. Validity and reliability of predicting maximum oxygen uptake via field tests in children and adolescents. *Pediatric Exercise Science*. 1991;3:250–5.

12. Chang RK, Gurvitz M, Rodriguez S, Hong E, Klitzner TS. Current practice of exercise stress testing among pediatric cardiology and pulmonology centers in the United States. *Pediatr Cardiol*. 2006;27(1):110–6.

13. Cumming GR, Everatt D, Hastman L. Bruce treadmill test in children: normal values in a clinic population. *Am J Cardiol*. 1978;41:69–75.

14. Dancey C, Reidy J. *Statistics without Maths for Psychology: Using SPSS for Windows*. London (United Kingdom): Prentice Hall, 2004:1–636.

15. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures? *J Clin Epidem*. 2006;59(10):1033–9.

16. Dencker M, Thorsson O, Karlsson MK, Lindén C, Wollmer P, Andersen LB. Maximal oxygen uptake versus maximal power output in children. *J Sports Sci*. 2008;26:1397–402.

17. Deurenberg P, van der Kooy K, Hautvast JG. The assessment of the body composition in the elderly by densitometry, anthropometry and bioelectrical impedance. *Basic Life Sci*. 1990;55:391–3.

18. Dubowy KO, Baden W, Bernitzki S, Peters B. A practical and transferable new protocol for treadmill testing of children and adults. *Cardiol Young*. 2008;18:615–23.

19. Foster C, Jackson AS, Pollock ML, et al. Generalized equations for predicting functional capacity from treadmill performance. *Am Heart J*. 1984;107(6):1229–34.

20. Haycock GB, Schwartz GJ, Wisotsky DH. Geometric method for measuring body surface area: a height-weight formula validated in infants, children, and adults. *J Pediatr*. 1978;93(1):62–6.

21. Hollenberg M, Ngo LH, Turner D, Tager IB. Treadmill exercise testing in an epidemiologic study of elderly subjects. *J Gerontol A Biol Sci Med Sci*. 1998;53:B259–67.

22. Johnston KN, Jenkins SC, Stick SM. Repeatability of peak oxygen uptake in children who are healthy. *Pediatr Phys Ther*. 2005;17(1):11–7.

23. Karila C, de Blic J, Waernessyckle S, Benoist MR, Scheinmann P. Cardiopulmonary exercise testing in children: an individualized protocol for workload increase. *Chest*. 2001;120:81–7.

24. Massicotte DR, Gauthier R, Markon P. Prediction of VO_{2max} from the running performance in children aged 10–17 years. *J Sports Med Phys Fitness*. 1985;25:10–7.

25. Medbø JI, Mamen A, Welde B, von Heimburg E, Strokke R. Examination of the Metamax I and II oxygen analysers during exercise studies in the laboratory. *Scand J Clin Lab Invest*. 2002;62:585–98.

26. Meyer T, Georg T, Becker C, Kindermann W. Reliability of gas exchange measurements from two different spiroergometry systems. *Int J Sports Med*. 2001;22:593–7.

27. Portney LG, Watkins MP. Statistical measures of reliability. In: Portney LG, Watkins MP, editors. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. Upper Saddle River (NJ): Pearson Education, Inc.; 2009. pp. 585–618.

28. Schönbeck Y, Talma H, van Dommelen P, et al. Increase in prevalence of overweight in Dutch children and adolescents: a comparison of nationwide growth studies in 1980, 1997 and 2009. *PLoS One*. 2011;6:e27608.

29. Shephard RJ, Allen C, Benade AJ, et al. The maximum oxygen intake. An international reference standard of cardiorespiratory fitness. *Bull World Health Organ*. 1968;38:757–64.

30. van der Cammen–van Zijp MH, van den Berg–Emons RJ, Willemsen SP, Stam HJ, Tibboel D, IJsselstijn H. Exercise capacity in Dutch children: new reference values for the Bruce treadmill protocol. *Scand J Med Sci Sports*. 2010;20:e130–6.

31. Wessel HU, Strasburger JF, Mitchell BM. New standards for the Bruce treadmill protocol in children and adolescent. *Pediatr Exerc Sci*. 2001;13:392–401.

32. Weststrate JA, Deurenberg P. Body composition in children: proposal for a method for calculating body fat percentage from total body density or skinfold-thickness measurements. *Am J Clin Nutr*. 1989;50:1104–15.