Reliability and accuracy of Cooper’s test in male long distance runners

J.R. Alvero-Cruz, M.A. Giráldez García, E.A. Carnero

Objectives: Endurance capacity can be assessed by field test such as Cooper’s test; however, reliability and accuracy are rarely reported in the literature. It was our aim to describe reliability and accuracy of Cooper’s test in long distance runners.

Method: Fifteen male long distance runners performed twice all-out Cooper’s test in a 400 m track. Total distance covered, maximum heart rate (HR) and rate of perceived exertion were recorded. Bias correction factor (Bc) was used to describe accuracy and the main dimensions of reliability were calculated by an intraclass correlation coefficient (ICC), effect size (ES) and agreement analysis.

Results: Accuracy for total distance and HR were relatively high (Cb = 0.994 and 0.956). Reliability for covered distance was as small as 1.7% (52.2 m) and ICC was 0.99; additionally, neither proportional nor systematical bias was detected in the agreement analysis.

Conclusions: All together, our results may confirm a good accuracy and reliability of Cooper’s test in an athlete long distance runners. Also, improvements or impairment lower than 52.2 m must not be associated with exercise training or retraining, since they are below the values of intra-subject reliability.

© 2016 Consejería de Turismo y Deporte de la Junta de Andalucía. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Palavras-chave:  
Atletas amadores  
Teste de campo resistência  
Fator de correção de viés  
Erro técnico de medição  
Análise de concordância  
Coeficiente de correlação intraclass  
Tamanho do efeito  

**Fiabilidade e precisão do teste de Cooper em corredores de longas distâncias do sexo masculino**

**Resumo**

Objetivo: A capacidade de resistência pode ser avaliada pelo teste de campo, tal como o teste de Cooper; no entanto, a fiabilidade e precisão são raramente relatados na literatura. O objetivo foi descrever a fiabilidade e precisão do teste de Cooper em corredores de longa distância.

Método: Quinze corredores de longa distância do sexo masculino realizaram teste de Cooper 2 vezes, numa faixa de 400 metros. Distância total percorrida, frequência cardíaca máxima (FC) e taxa de esforço percebido foram registradas. Fator de correção do viés (BC) foi usado para descrever a precisão e as principais dimensões de fiabilidade foram calculados por meio do coeficiente de correlação de intraclass (ICC), tamanho do efeito (ES) e análise de concordância.

Resultados: A precisão da distância total e frequência cardíaca eram relativamente altas (Cb = 0.994 e 0.956). Fiabilidade para o curso era tão pequena quanto 1.7% (52.2 metros) e ICC de 0.99, além disso, uma vez que nem viés proporcional, nem sistemático foram detetados através da análise de jogo.

Conclusões: Os nossos resultados podem confirmar uma boa precisão e fiabilidade do teste de Cooper em corredores de longa distância amadores. Além disso, melhorias ou prejuízo menor do que 52.2 metros não devem ser associados a treinamento físico ou destreinamento, uma vez que estão abaixo dos valores de fiabilidade intrassujeitos.

© 2016 Consejería de Turismo y Deporte de la Junta de Andalucía. Publicado por Elsevier España, S.L.U. Este é um artigo Open Access sob uma licença CC BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/4.0/).

**Introduction**

Maximum oxygen uptake (VO2max), lactate thresholds and running economy have been widely used to assess endurance and aerobic capacity in middle and long distance runners, and all related to athletic performance. However, these variables are time consuming and expensive in field settings still; indirect tests can be utilized to substitute these latter assessments. The utility of a test depends on its validity, accuracy and reliability (reproducibility). Validity can be assumed if a test represents accurately those features of the phenomena, which are aimed to describe, explain or theorise.

Regarding accuracy, this is the degree of a test to measure the true value. Finally, reliability informs about reproducibility of a test and a procedure of repeated measures is used in order to calculate repeatability; so we can consider reliability as the degree to which an assessment tool produces stable and consistent results (also known as test–retest reliability). Both low reliability and accuracy may limit applicability and utility of field performance tests.

However, utility of field tests has commonly relied on construct validity, usually associated with the capacity of the test to estimate or be associated with laboratory variables or clinical tests. In this sense, one of the most studied physiological constructs is VO2max, which determines the maximum aerobic capacity and should be related with endurance and long-term performance. Thus, several field tests have been created in order to obtain a valid and reliable estimation of VO2max. One of the first tests developed to estimate VO2max was Cooper’s test, which is a simple time limit single-stage test, where athletes need to cover as many meters as possible during a 12-min all-out test. The VO2max estimated from Cooper and a multistage shuttle run tests has been strongly correlated in young healthy adults, which may confer a good concurrence at least for this population. The same study showed a good reliability (Φ: 0.96) and acceptable systematic error of 4.3% for maximal oxygen uptake prediction. However, the Cooper’s test accuracy has not been still reported to date. Also, there are a lack of data of reliability and accuracy data in athletes.

Since, there is a lack of knowledge about the reproducibility (test–retest reliability) characteristics of field tests to estimate endurance capacity such as Cooper’s test in long distance runners, it was our aim to analyze the reliability and accuracy of Cooper’s test on amateur long distance runners over two repeated measures (test–retest).

**Method**

**Subjects**

Fifteen adult male amateur athletes (34.5 ± 1.9 years, and 3.7 ± 4.6 years of training) volunteered to participate in the study. All athletes were informed of the study characteristics, procedures and risks; afterwards a signed informed consent was obtained from those who decided to be enrolled. The Ethical Review Institutional Board (IRB) at the University of Malaga approved the research protocol.

**Experimental procedures**

Test–retest approach was used by repeating Cooper’s test twice in a period of 48 h. Reliability analysis was carried out in all variables obtained from the Cooper test such as distance, heart rate (HR) at the end of the test and the rate of perceived exertion (RPE). Two Cooper’s tests split by 48 h were carried out in a synthetic track of 400 m, and under similar meteorological conditions. Every day athletes followed thoroughly the same protocol: firstly, a 15-min running warm-up was performed at between 50 and 70% of the theoretical maximal HR (220-Age). Then, the original Cooper’s test was executed; briefly, athletes were asked to run all-out during 12-min along the inner lane of the track; immediately afterwards a member of research team recorded the distance in meters by placing a mark exactly in the point where every athlete stood still. Also, the HR at the end of test was recorded by using a HR monitor Polar RS300X (Polar Electro, Finland), and the RPE using the 0–10 Borg scale was individually asked to each participant.

**Statistical analysis**

The accuracy of total distance in Cooper’s test, maximal HR and RPE were calculated by bias correction factor (Cb) from concordance correlation coefficient analysis. Absolute reliability was reported as the mean differences, coefficient of variation (CV),
Table 1: Anthropometric and training variables of the sample.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (kg)</td>
<td>67.3 ± 10.7</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>171.0 ± 6.8</td>
</tr>
<tr>
<td>Age (years)</td>
<td>34.5 ± 1.9</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>22.9 ± 1.5</td>
</tr>
<tr>
<td>Training time (years)</td>
<td>3.7 ± 4.6</td>
</tr>
<tr>
<td>Km/week (km)</td>
<td>44.8 ± 9.8</td>
</tr>
</tbody>
</table>

\( \sqrt{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 / n\right)} \), the standard error of the mean (SEM) and the effect size (ES) using the d coefficient of Cohen. For this study, an ICC < 0.50 was considered fair; from 0.50 to 0.75 was considered good and >0.75 excellent. Also, Cohen’s d ES of 0.20 was considered small, 0.50 medium, and 0.80 large.

No significant differences were found between test 1 and 2 either for total distance or HR. Additionally, our ICC results from test–retest data indicated that Cooper’s test had a very good reliability for covered distance and HR (Table 2). Regarding RPE, we observed a good ICC, although a significant difference was found between RPE in the first and second test \((P < 0.001, \text{Table 2})\).

Agreement analysis from the Bland–Altman plots did not showed systematic error for both, distance \((\text{difference} = 20.5 \text{ m}, P > 0.05)\) or maximal HR \((\text{difference} = −1.1 \text{ bpm}, P > 0.05)\), neither proportional bias as confirmed by Kendall’s Tau rank correlation coefficient between differences and mean of measurements (Fig. 1).

Discussion

The aim of this study was to perform a preliminary reliability and accuracy of the Cooper’s test in amateur long-distance runners. Our data support a good reliability as suggested previously by other authors, who studied the reliability of Cooper’s test in non-athletic samples.\(^6\) In spite of small differences between the two trials, CV of Cooper’s test remained still around 52.2 m, although in relative units it was as low as 1.7%. This moderately high CV could be explained by the great heterogeneity of the athletic performance of the sample \((\text{range: 2350–3520 m trial 1 and 2275–3540 m trial 2})\), so the same absolute distance may represent similar percentages for high and low extremes in performance. In spite of the limitation, this may offer better generalization of our results since they included a larger range of performances and may highlight the bias of reliability data from a previous study where a more homogenous sample than ours was analyzed.\(^5\) Moreover, the ES of the differences was as low as 0.059 and the non-significant difference on covered distances between trials may indicate the good repeatability of this test.

Table 2: Relative and absolute reliability of Cooper’s test variables.

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Distance 1 (m)</th>
<th>Distance 2 (m)</th>
<th>HR1 (bpm)</th>
<th>HR2 (bpm)</th>
<th>RPE1</th>
<th>RPE2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ± SD</td>
<td>3026 ± 330</td>
<td>3047 ± 359</td>
<td>182 ± 7.3</td>
<td>183 ± 5.7</td>
<td>8.7 ± 0.6</td>
<td>9.5 ± 0.5</td>
</tr>
<tr>
<td>Mean diff (95% CI)</td>
<td>20.46 (−20.22 to 61.15)</td>
<td>0.99 (0.96–0.99)</td>
<td>1.13 (−0.69 to 2.93)</td>
<td>0.93 (0.80–0.98)</td>
<td>0.8 (0.48–1.11)</td>
<td></td>
</tr>
<tr>
<td>ICC (95% CI)</td>
<td>0.68 (0.50–0.89)</td>
<td>0.7 (0.5–0.95)</td>
<td>0.68 (0.05–0.89)</td>
<td>0.7 (0.5–0.95)</td>
<td>0.1447</td>
<td></td>
</tr>
<tr>
<td>CV (CV %)</td>
<td>52.2 (1.7%)</td>
<td>52.2 (1.7%)</td>
<td>2.4 (1.3%)</td>
<td>0.8387</td>
<td>0.173</td>
<td>1.405</td>
</tr>
<tr>
<td>SEM</td>
<td>18.97</td>
<td>18.97</td>
<td>0.173</td>
<td>0.173</td>
<td>0.173</td>
<td>1.405</td>
</tr>
<tr>
<td>Cohen’s d</td>
<td>0.059</td>
<td>0.059</td>
<td>0.059</td>
<td>0.059</td>
<td>0.059</td>
<td>0.059</td>
</tr>
</tbody>
</table>

Data in the table are from two repeated all–out Cooper’s test. 1 and 2 subscripts indicate first and second Cooper’s test respectively. HR, maximal heart rate during the last minute of the test; SD, standard deviation; Mean diff, mean difference between first and second test; ICC, interval of confidence; IC, intraclass correlation coefficient; CV, coefficient of variation \((CV(\text{original units}) = \sqrt{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 / n\right)} / \text{mean} \times 100\); SEM, standard error of the mean; RPE, rate of perceived exertion \((\text{scale from 0 to 10})\).

\(^1\) P < 0.001, for paired sample T-test.
Firstly, these results may be helpful for coaches and scientists when prescribing training load, reporting VO_{2max} changes or predicting performance in order to interpret the variability of their outcomes. On the other hand, researchers could use these data in order to calculate sample size. This study does not lack of limitations, and our results could be biased by the intensity of test, so it can be argued that the athletes did not exercise at maximum or same effort in both trials. By using HR, the intensity of aerobic exercise test may be easily confirmed. In this study, all participants reached theoretical maximal HR values as predicted from age, which may suggest that both trials were performed all-1 out. In relation with heart rate reliability, it was also observed a CV was also observed among 4 and 3.1%, a low effect size of the difference (0.17), as well as very low absolute reliability for the maximal HR (1.13 bpm); all together these results suggest that trials 1 and 2 were similar in intensity. Additionally, RPE is a recognized marker of intensity and homeostatic disturbance during exercise and it is usually monitored during exercise tests to complement other dimensions of intensity. Garcin analyzed the reliability of the HR and RPE in progressive and constant intensity exercises, concluding that these variables are reliable and replicable in these exercises. Nevertheless, our results did not confirm this latter evidence and RPE had a low reliability as confirmed by the very large ES found (1.4). A plausible reason for this disagreement may be related with the poor experience of athletes in using this variable.

In conclusion our results showed that the Cooper’s test is highly reliable when repeated after 48 h as confirmed by HR and distance data. This study provided support for the Cooper’s test as an accurate and reliable test to assess performance in a sample of amateur long-distance runners. Nonetheless, more studies are it must be necessary in order to validate performance-related constructs with Cooper’s test to confirm its utility as training tool in field settings.

Conflict of interest

The authors declare to have no conflict of interest.

Acknowledgements

We gratefully acknowledge the participants who dedicated their time to collaborate in this study, especially to coaches Juan Vázquez Sánchez and Daniel Pérez Martínez.

References