CONCURRENT VALIDITY
AND INTER-RATER RELIABILITY
OF HAND-HELD MEASUREMENTS
OF MAXIMAL SPRINT SPEED

Bartosz Zając1 A,B,C,D,E,F,G, Paulina Katarzyna Gaj2 A,B,C,D,E,F,G, and Jakub Mateusz Zięba3 A,B,D

1 Laboratory of Functional Diagnostics, Central Scientific and Research Laboratory, University of Physical Education in Kraków, Poland
2 Cracow Academy of Track-and-Field Sports, Kraków, Poland
3 University of Physical Education in Kraków, Poland

Keywords: concurrent validity, inter-rater reliability, hand-held timing, maximal sprint speed

Abstract:
Background: Maximal sprinting speed (MSS) is an essential component of success in many sports. Currently, many systems are used to accurately evaluate athletes’ MSS, including laser or radar guns, single- or dual-beam photocells, high-speed cameras, and high-frequency global positioning systems. However, the cost of these devices may be an obstacle to their implementation into practice. The least expensive but most likely less accurate alternative method of MSS evaluation is the hand-held time measurement of a 30m flying-start sprint. Therefore, the aim of the study was to assess the concurrent validity and inter-rater reliability of the hand-held method of MSS measurement.

Materials and Methods: The study involved 3 experienced raters and 18 amateur runners. Runners performed 2-3 trials of the 30m maximal flying-start sprint. In total, 40 observations were collected. Each sprint time was measured simultaneously by raters using a hand-held stopwatch and an electronic timing system. Criterion validity (hand-held vs. electronic timing) was assessed using linear regression analysis. Inter-rater reliability between hand-held timers was evaluated using interclass correlation coefficients (ICCs), standard error of measurement (SEM), and minimal detectable change (MDC).

Results: Results showed that single and average hand-held methods are affected by -0.17 to -0.07m·s⁻¹ (-2.5 to -1.6%) and -0.12m·s⁻¹ (-1.7%) errors, respectively. Linear regression analysis parameters (free parameter not statistically significant, directional coefficient 0.994-1.057, standard error of estimation 0.073-0.125, R² 0.981-0.994) indicated statistically excellent absolute agreement between a hand-held (single and average) and electronic timing. ICCs of 0.980-0.994, SEM of 0.12m·s⁻¹ (1.87%), and MDC of 0.34m·s⁻¹ (5.18%) indicated statistically excellent absolute agreement and consistency for single and average measurements between hand-held timers.

Conclusion: The proposed manual method of MSS measurement underestimates athletes’ speed performance. Moreover, the hand-held 30m flying-start sprint time measurement is affected by a 2% error, and a minimum 5% time change in an individual athlete demonstrates that the change is not simply attributable to measurement error.

Introduction
Maximal sprinting speed (MSS) is an essential component of success in many sports [1–6]. Therefore, the best training methods to improve sprint performance are of interest to many strength and conditioning coaches and scientists [7–11]. Valid and reliable MSS measurements are critical for the effective monitoring of performance and depend on standardized procedures and the accuracy of devices.
Currently, many systems are used to accurately evaluate athletes' MSS, including laser or radar guns, single- or dual-beam photocells, high-speed cameras, and high-frequency global positioning systems [12]. However, the cost of these devices is an obstacle to their implementation into practice. The least expensive but most likely less accurate alternative method to evaluate MSS is the hand-held time measurement of a 30m flying-start (FS) sprint, which is based on the assumption that athletes attain maximal speed at 30 to 40m of the sprint and maintain it at a relatively constant level up to 60 to 70m [13–15].

There are some reports on the reliability and validity of hand-held sprint time measurements performed from a standing-start (SS) position [16–19]. Fry and Kraemer [16] suggested adding 0.20s to hand times to represent electronic times for the 40yd (36.6m) sprint. International rules stipulate that 0.24s be added to any hand-timed mark in the 100 or 200-m events when converting between the two [17]. Mann et al. [18] and Mayhew et al. [19] showed that experienced hand-held raters produce a significantly faster (0.22±0.07s, 0.31±0.07s, respectively) 40yd sprint time than electronic timing. In addition, the authors of this study reported statistically high reliability between experienced hand-held raters and electronic timing (>0.956 interclass correlation coefficient). However, the specificity of the flying-start (FS) sprint time measurement is different from that of SS. Raters assessing the FS sprint need to carry out observation from a certain distance and at a certain angle from the place of the measurement start and end. Moreover, the measurement is started when a runner moves, which makes it more difficult. Interestingly, there is only one study [20] in which the reliability and validity of hand-held FS sprint timing have been assessed. The authors of this research, using multiple-split timers, stood in the center of a standard 400-m track and recorded each 25-m split of a 200-m sprint trial, reporting a mean error of -0.16±0.19s and reliability >0.895 interclass correlation coefficient. It is necessary to highlight that testing the protocol set-up utilized by the authors of this study precludes raters focusing on 1 section assessment.

To the best of our knowledge, there are no studies evaluating the reliability and validity of the hand-held method of maximal sprinting speed measurement. Information about expected measurement errors may be valuable for practitioners who do not have access to more accurate measurement methods. This data may be useful, for example, in the interpretation of the effectiveness of sprint training. Therefore, the purpose of the study was to assess the concurrent validity and inter-rater reliability of the hand-held method of maximal sprint speed measurement.

**Material and Methods**

**Participants**

The study involved 3 raters recruited from track-and-field coaches from Kraków, Poland. The criteria for inclusion into the group of raters were (i) having a coaching certificate issued by the Polish Athletic Association and (ii) at least 3 years of experience in measurements of sprint and agility performance of athletes. The criterion for exclusion from the group of raters was visual impairment. Furthermore, 18 amateur runners of both sexes (age: 39±9.3 years, body height 172±8.1cm, body mass: 65±10.4kg) recruited from a local sports club participated in the study. The inclusion criteria for the group of runners were (i) age above 18 years, (ii) having a current certificate from a sports medicine doctor regarding the ability to practice track-and-field. The exclusion criterion for the group of runners was acute injuries making it impossible to participate in the tests. Informed consent was obtained from all subjects involved in the study.

**Study design**

All measurements were performed on the same day between 2:00 p.m. and 5:00 p.m. on an outdoor athletics track at an ambient temperature of 29±1°C. Runners performed 2-3 trials (30-m maximal flying-start running sprint) separated by at least a 5-minute rest interval. In total, 40 observations were performed. The sample size was calculated by using an online sample size calculator [21] (input parameters: minimum acceptable reliability of 0.8, expected reliability of 0.9, significance level set at 0.05, power of 0.8, four raters, and expected dropout of 10%). Before trials, the runners performed individual warm-ups that included jogging, drills, and various stretching exercises. Each sprint time was simultaneously measured to the nearest 0.01s by raters using Onstart 710 hand-held stopwatches (Geonaute, China) and the Witty Gate electronic timing system (Microgate S.r.l, Italy). Photocells were placed 80cm above the ground [22]. Raters were instructed to start and stop their stopwatches when the upper body crossed the photocells and changed the position after each measurement. Prior to the trials, raters were familiarized with the characteristics of the stopwatch (i.e. stiffness of the start button and the depth of its depression needed to activate the watch). A diagram of the testing protocol is presented in Figure 1. All procedures were carried out in accordance with the 1964 Declaration of Helsinki and its subsequent amendments. The testing protocol was approved by the Bioethics Committee at the Regional Medical Chamber in Kraków (No. 249/KBL/OIL/2021).
Concurrent Validity and Inter-Rater Reliability...

Statistical analysis

Statistical analysis was performed using PQStat v. 1.8.4 software (PQStat Software Company, Poland). Concurrent validity (hand-held vs. electronic timing) was evaluated using linear regression analysis [23]. Inter-rater reliability between hand-held timers was assessed using (ICCs) 2.1, 2.k, 3.1, 3.k [24], standard error of measurement [25], and minimal detectable change [25]. The ICCs were interpreted according to Portney and Watkins [26] as poor (below 0.50), moderate (from 0.50 to 0.75), good (from 0.75 to 0.90), and excellent (above 0.90). The probability of Type I error below 0.05 was adopted as the level of significance.

Results

Compared to photocells, hand-held measurements of maximal sprint speed (MSS) underestimate athletes’ performance. However, linear regression analysis parameters indicated statistically excellent absolute agreement between the hand-held (single, average) and electronic methods. Detailed numerical data are presented in Table 1.

Table 1. Concurrent validity of hand-held timing

<table>
<thead>
<tr>
<th></th>
<th>Average velocity [m·s⁻¹] mean (SD) min-max</th>
<th>Mean error [m·s⁻¹] (95%CI)</th>
<th>Mean error [%] (95%CI)</th>
<th>β₀ (SE) p-value</th>
<th>β₁ (SE) p-value</th>
<th>ε</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT₁</td>
<td>6.56 (0.89) 5.34-8.11</td>
<td>-0.11 (-0.15-(-0.07))</td>
<td>-1.6 (-2.2-(-1.1))</td>
<td>0.149 (0.147)</td>
<td>0.315</td>
<td>0.125</td>
<td>0.981</td>
</tr>
<tr>
<td>HT₂</td>
<td>6.50 (0.84) 5.22-7.89</td>
<td>-0.17 (-0.21-(-0.13))</td>
<td>-2.5 (-3.1-(-2.0))</td>
<td>-0.198 (0.154)</td>
<td>0.206</td>
<td>0.125</td>
<td>0.981</td>
</tr>
<tr>
<td>HT₃</td>
<td>6.60 (0.89) 5.38-8.11</td>
<td>-0.07 (-0.11-(-0.04))</td>
<td>-1.1 (-1.6-(-0.6))</td>
<td>0.050 (0.131)</td>
<td>0.703</td>
<td>0.110</td>
<td>0.985</td>
</tr>
<tr>
<td>HTA</td>
<td>6.55 (0.87) 5.31-7.97</td>
<td>-0.12 (-0.14-(-0.09))</td>
<td>-1.7 (-2.1-(-1.4))</td>
<td>-0.071 (0.088)</td>
<td>0.425</td>
<td>0.073</td>
<td>0.994</td>
</tr>
<tr>
<td>ET</td>
<td>6.67 (0.90) 5.31-8.15</td>
<td>ref.</td>
<td>ref.</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
</tr>
</tbody>
</table>

HT₁,₂ – hand-held timers, HT₃ – average of hand-held timers, ET – electronic timing. Mean error – difference (HT-ET), SD – standard deviation, ICC – interclass correlation coefficient (HT vs. ET), 95%CI – 95% confidence interval, β₀ – free parameter, β₁ – directional coefficient, SE – standard error, ε – standard error of estimation, R² – determination coefficient, ref. – reference value, NA – not applicable

Interclass correlation coefficients (ICCs) exhibited statistically excellent agreement and consistency for single and average measurements between hand-held timers. Detailed numerical data regarding inter-rater reliability are presented in Table 2.

Table 2. Inter-rater reliability between hand-held timers

<table>
<thead>
<tr>
<th></th>
<th>ICC (2,1) (95%CI)</th>
<th>ICC (2,k) (95%CI)</th>
<th>ICC (3,1) (95%CI)</th>
<th>ICC (3,k) (95%CI)</th>
<th>SEM [m·s⁻¹]</th>
<th>SEM [%]</th>
<th>MDCₑ [m·s⁻¹]</th>
<th>MDCₚₑ [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.980 (0.965-0.989)</td>
<td>0.993 (0.988-0.996)</td>
<td>0.983 (0.971-0.990)</td>
<td>0.994 (0.990-0.997)</td>
<td>0.12</td>
<td>1.87</td>
<td>0.34</td>
<td>5.18</td>
</tr>
</tbody>
</table>

ICC – interclass correlation coefficient, 95%CI – 95% confidence interval, SEM – standard error of measurement, MDC – minimal detectable change
**Discussion**

Our research is the first to evaluate concurrent validity and inter-rater reliability of the hand-held method of maximal sprint speed (MSS) measurement. The most important finding of the present study was that maximal sprint speed measurement using the testing protocol proposed by the authors of this study underestimates athletes’ speed performance. Moreover, in hand-held time measurement of a 30m flying-start sprint, a 2% error and at least a 5% time change are added for an individual athlete, which confirms that the change is not simply attributable to measurement error.

In our study, we noted that MSS estimation using single and average hand-held time measurement of the 30m flying-start sprint is affected by a mean error from -2.5 to -1.1% (-0.17 to -0.07m·s⁻¹) and -1.7% (-0.12·ms⁻¹) error compared to electronic timing with split time for 30m. This is very interesting because the authors, who assessed the validity of hand-held time measurement of a 40-ya (36.6m) standing-start sprint obtained opposite results [18,19,27]. Mann et al. [18] demonstrated that experienced and novice hand-held timers produce a significantly faster sprint time than electronic timing (-0.22±0.07s and -0.26±0.08s, respectively). Mayhew et al. [19] showed that hand timing is significantly faster than electronic, producing a mean error of -6.0±1.3% (-0.31±0.07s). Moore et al. [27] noted a mean difference of -0.08±0.16s between the average of 3 hand timers and electronic timing. The most likely explanation for the differences between hand and electronic standing-start sprint timing is simple reaction time latency, which widely varies and ranges from 0.23s [28,29] to nearly 0.40s [30]. Moreover, Haugen et al. [31] noted that electronic timing may actually start 0.04s before visual hand movement from a starting pad can be detected by video analysis. However, it is unclear why flying-start hand-held timing produces reverse errors. It is probable that flying-start sprint timers anticipate the moment when athletes cross the start photocell [32], whereas standing-start sprint timers are “surprised” by the starting runner. Nonetheless, this supposition is contrary to that presented by Hetzler et al. [20]. In their study, which used multiple-split timing, the researchers reported an absolute error of -0.16±0.19s. It should be noted, however, that the testing protocol utilized by the authors of this study provided recordings for each 25m split of the 200m sprint trial, which precluded raters to be focused on 1 section.

MSS measurement based on the average velocity of a 30m flying-start running sprint is vitiated by some error related to running speed instability during sprinting. Speed instability results from the shape of the sprinting velocity curve on which one can distinguish 3 phases: acceleration, constant velocity, and deceleration [33]. Moreover, running velocity fluctuates during each step of a sprint due to anterior-posterior forces [34]. It is therefore necessary to expect an increase in the accuracy of measurement of maximal sprint speed (and higher velocities) along with an increase in the sampling rate used in the method, as confirmed by the results of empirical research. Alphin et al. [35] demonstrated that MSS measurement using a 30m flying start sprint and high-frequency (10Hz) global positioning system produces a mean bias of +0.44m·s⁻¹ and mean absolute percent error of +5.5±2.6% compared to the use of timing gates (split times of 10m). Zabaloy et al. [36] showed mean differences of +0.082 and +0.036m·s⁻¹ when MSS was measured using a radar gun (sampling frequency of 47Hz) compared to timing gates (split times of 10m and 5m, respectively).

ICCs of 0.980-0.994, standard error of measurement (SEM) of 0.12m·s⁻¹ (1.87%), and minimal detectable change (MDC) of 0.34m·s⁻¹ (5.03%) indicate statistically excellent inter-rater reliability (absolute agreement and consistency) between single and average hand-held timing. In previous studies (18,20), evaluation of hand-held sprint time measurement reliability yielded similar results to those obtained in our study. Mann et al. [18] reported an ICC of 0.972 for experienced timers and ICC 0.972 for novice timers. Hetzler et al. [20] showed an ICC of 0.90 and SEM of 0.12s for multi-split timers. Other methods to measure MSS have also demonstrated similar reliability. Johnston et al. [37] reported ICC of 0.97 for a 10Hz global positioning system. Talukdar et al. [38] demonstrated intra-day (ICC: 0.85, 95%CI: 0.70-0.93) and inter-day (ICC: 0.98, 95%CI: 0.98-0.99) reliability of maximal velocity assessment using a radar gun. When interpreting indicators of tool reliability, one should bear their purpose in mind. In assessing athletes’ speed potential and/or monitoring speed training effectiveness, higher measurement accuracy is required than that of hand-held measurements, which has been reported in our study and those carried out by other authors. For example, the time difference between the 1st and 6th finalists of the 60-meter women’s World Athletic Indoor Championship in 2022 was 1.15% [39]. If the timing had been performed by a stopwatch, the first 6 athletes would have had a time difference within the standard range for measurement error. Furthermore, using hand timing, a minimum 5.18% speed change in an individual athlete ensures (with 95% probability) that the change is not simply attributable to measurement error. This accuracy is insufficient for efficient training monitoring because a 5% speed performance improvement is quite a challenge [40].

The limitation of this study is to use timing gates with split time for 30m as a concurrent criterion. The protocol set-up underestimated the real maximal sprint speeds and overstated the method’s validity. Therefore, a better so-
lution would be to use a method based on a higher sampling rate as a concurrent criterion, such as, for instance, timing gates with split time for 5m or a radar/laser gun. The fastest runner in this study achieved an average velocity of 8.15m·s⁻¹. It is possible that faster runners (running velocity 9-11m·s⁻¹ is not uncommon among athletes) may be more difficult to accurately time using a hand-held stopwatch. Moreover, a heterogeneous (in terms of speed performance) group was included, and consequently, measurement consistency may be overestimated.

**Conclusion**

The proposed hand-held measurement method of maximal sprint speed underestimates athletes’ speed performance. Moreover, this method has a 2% error, and a minimum 5% speed change in an individual athlete, confirming that the change is not simply attributable to the measurement error.

**Founding**

This study was performed within the framework of the programme provided by the Minister of Science and Higher Education under the name ‘Regional Initiative for Perfection’ within the years 2019–2022, project No. 022/RID/2018/19, supported by the National Science Center in Poland.

**Institutional Review Board Statement:** All procedures were carried out in accordance with the 1964 Declaration of Helsinki and its subsequent amendments. Consent to perform testing was provided by the Bioethics Committee at the Regional Medical Chamber in Kraków (No. 249/KBL/OIL/2021).

**Informed consent statement:** Informed consent was obtained from all subjects involved in the study.

**Data availability statement:** The data presented in this study are available on request from the corresponding author.

**Conflict of interest:** The authors declare no conflict of interest.
References:


Concurrent Validity and Inter-Rater Reliability... 


Cited:

Author for correspondence
Bartosz Zajac
ORCID: 0000-0001-5398-4224
Affiliation: Laboratory of Functional Diagnostics, Central Scientific and Research Library, University of Physical Education in Kraków, Poland
Email: bartosz.zajac@awf.krakow.pl
Phone: +48 508-116-286