LITHUANIAN SPORTS UNIVERSITY

NORDPLUS

Digitalization in the coaching process across Nordic-Baltic countries (NPHZ-2022/10033)

Reliability and validity of tracking, monitoring and testing systems

Marco PERNIGONI, PhD

2024-11-19

Before we start

About the lecturer

Marco Pernigoni, MSc

marco.pernigoni@lsu.lt

- **2018** - Bachelor's degree in Exercise, Sport and Health Sciences (University of Milan, Italy)
- **2020** - Master's Degree in Individual and Team Sport Science (University of Milan, Italy)
- **2019/20** - Assistant S&C coach at Blu Basket 1971 Treviglio (Italian Men's Basketball Division II)
- **Oct 2020-present:** PhD student / Research assistant / Lecturer (Lithuanian Sports University)
  Topic: "Investigating recovery strategies to optimize performance, sleep and well-being in basketball"
- **Oct 2021 – May 2022:** Visiting PhD student (Örebro University, Sweden)
- **Jun – Jul 2023:** Visiting PhD student (University of Rome "Foro Italico", Italy)

In this lecture

- Tracking systems and monitoring tools: What are they?
- Validity and reliability: Foundational concepts
- A brief note: Factors affecting validity and reliability

Tracking systems and monitoring tools: What are they?

Tracking systems and monitoring tools

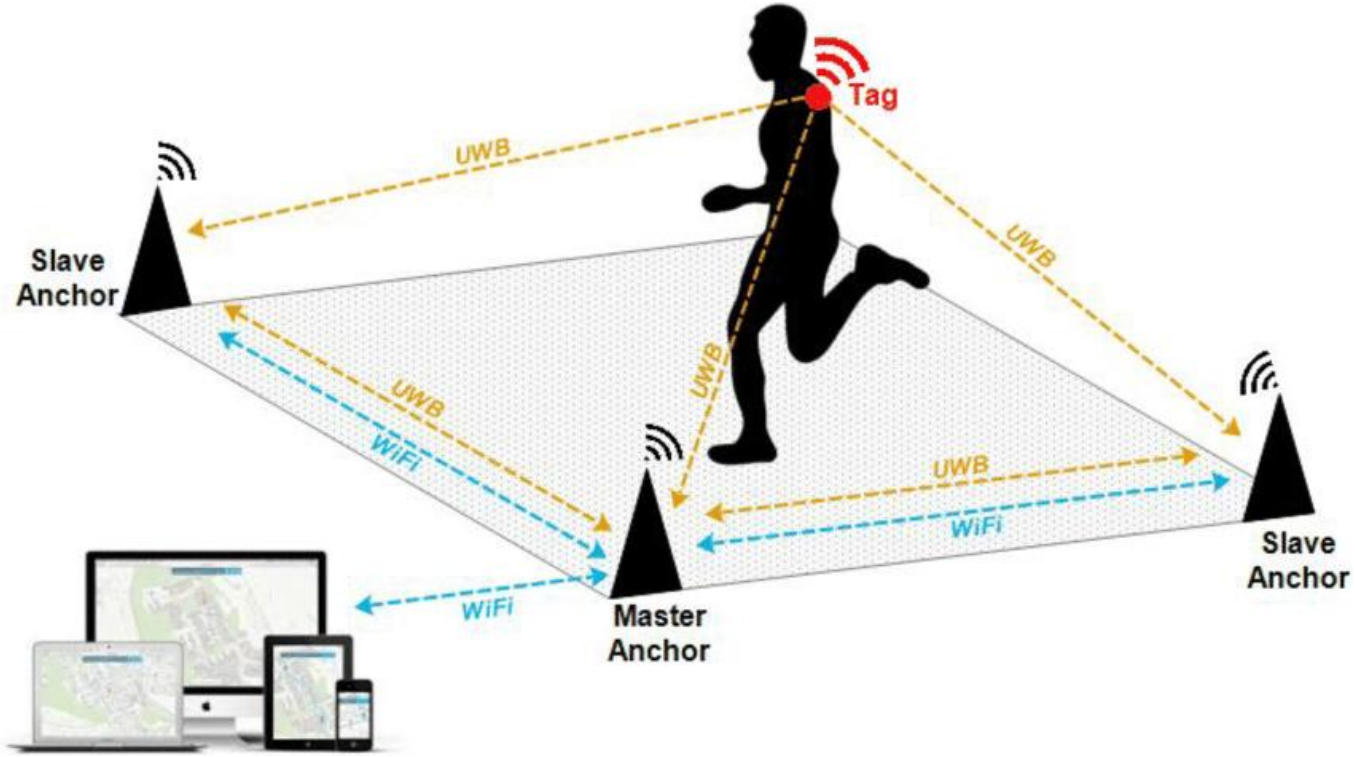
Tracking systems and monitoring tools

Technologies / methodologies / tools used to monitor and record movements, actions and physiological data, etc. before, during and/or following exercise

Common applications of tracking systems:

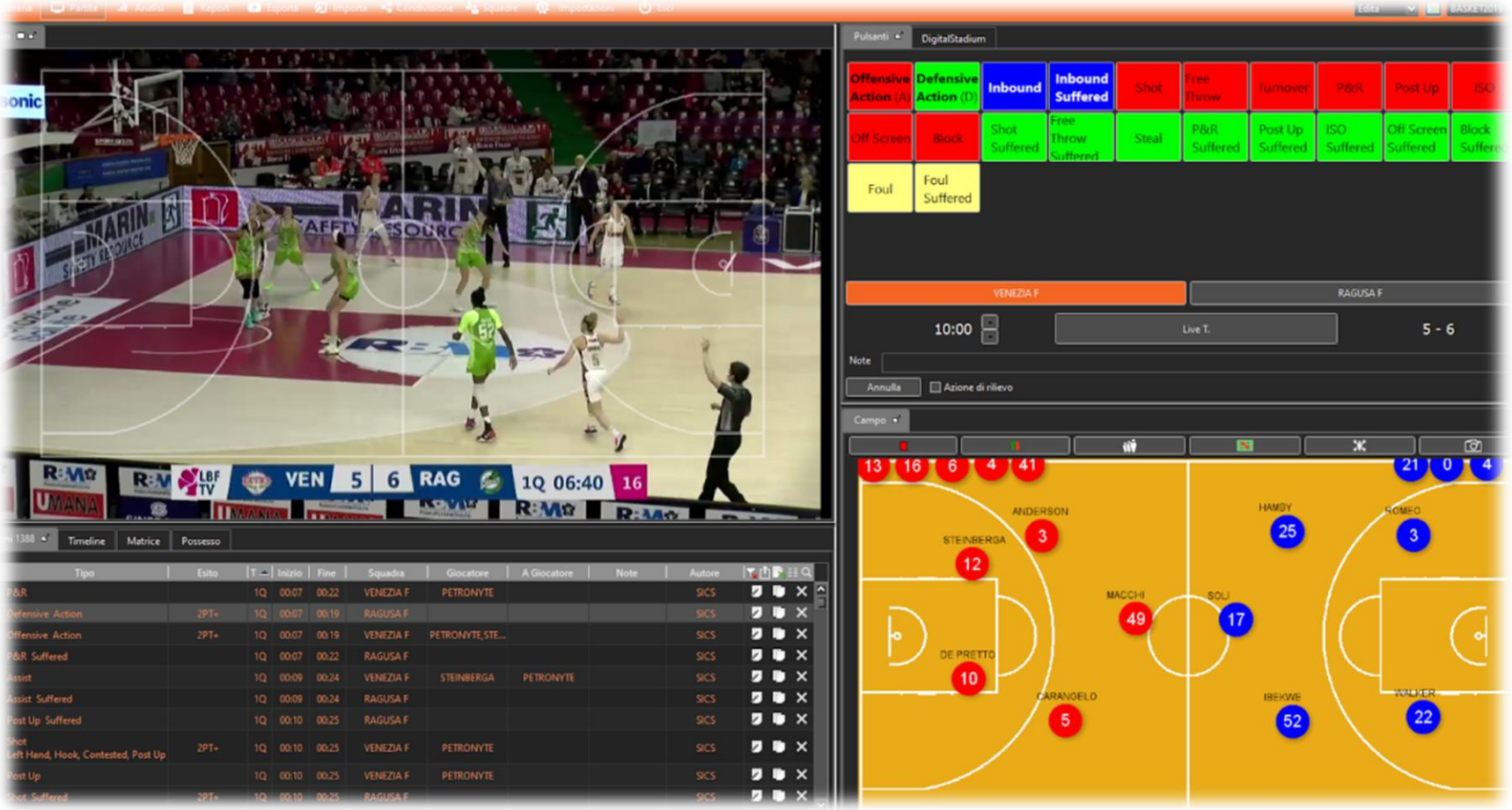
- Physical / technical / tactical performance analysis
- Load and fatigue monitoring
- Training planning
- Injury prevention

Tracking systems and monitoring tools



Tracking systems and monitoring tools

Tracking systems and monitoring tools



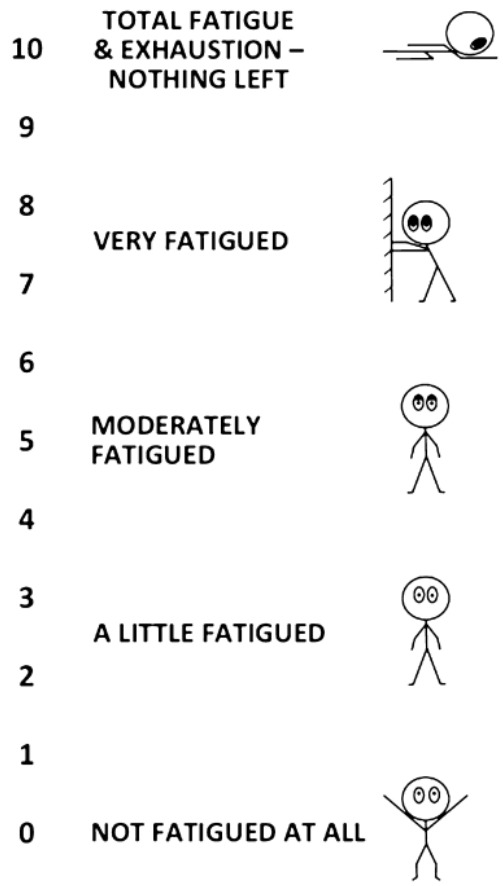
Tracking systems and monitoring tools

Tracking systems and monitoring tools

OPTOJUMP next

Tracking systems and monitoring tools

Ratings of perceived exertion (RPE)	Total quality recovery (TQR)
6	6
7 Very, very light	7 Very, very poor recovery
8	8
9 Very light	9 Very poor recovery
10	10
11 Fairly light	11 Poor recovery
12	12
13 Somewhat hard	13 Reasonable recovery
14	14
15 Hard	15 Good recovery
16	16
17 Very hard	17 Very good recovery
18	18
19 Very, very hard	19 Very, very good recovery
20	20




and many
More...

Testing procedures

Testing
=
The objective core of the athlete evaluation process

Tests
=
Procedures for assessing the athlete's ability in a particular area or skill

Purpose of testing

- **Assessment of Athletic Talent**
  Does the athlete have potential to compete?
- **Identification of Physical Abilities in Need of Improvement**
  Which qualities need to be central in the training plan?
- **Goal setting**
  What level of performance is desired in the future, and what is the current condition?

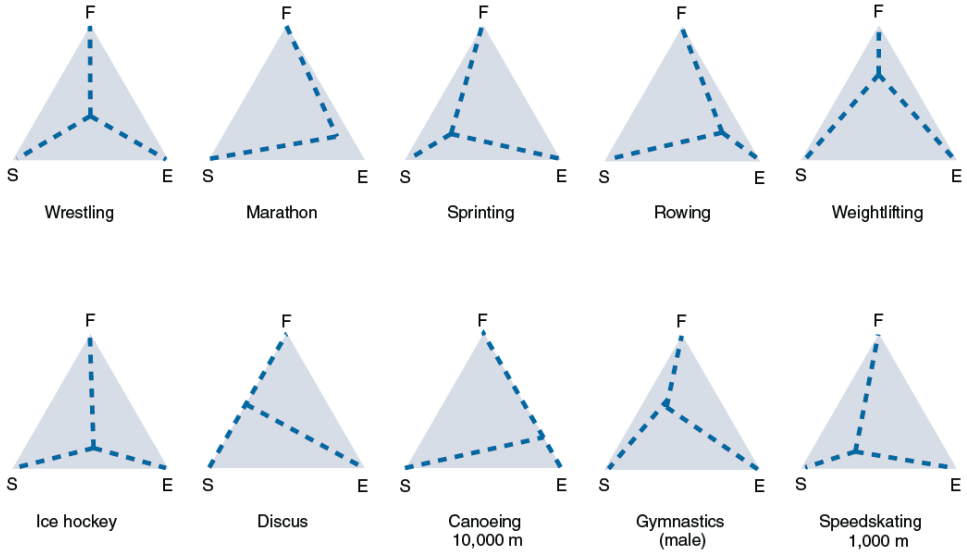


FIGURE 10.2 The dominant composition among the biomotor abilities of various sports.
F = strength or force; S = speed; E = endurance.

Utility of tracking, monitoring and testing

We are lucky to have so many «arrows in our quiver», that we can use to support the coaching process

Utility of tracking, monitoring and testing

Tracking systems, monitoring tools and testing procedures are resources that can be utilized in the coaching process, and provide valuable information

However…

We must ensure their correct implementation

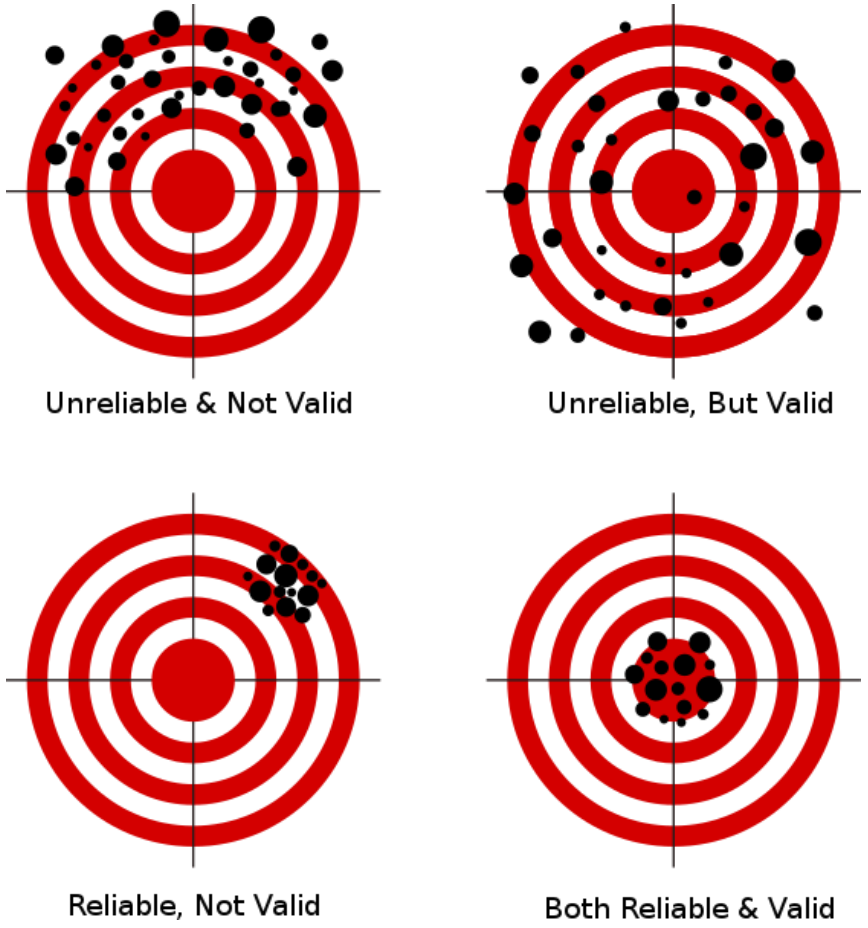
In other words…

We must make sure that the tools we use are **accurate**, **reliable** and provide **relevant** information

Validity and reliability: Foundational concepts

Validity and reliability

- **Validity** → the degree to which a tool or testing procedure measures what it is supposed to measure
- **Reliability** → the degree of consistency or repeatability of a tool or testing procedure



Validity

- **Construct validity** → ability of a tool / testing procedure to represent the underlying construct (see previous slide)
- **Criterion validity** → the extent to which scores from a tool / testing procedure are associated with other measures of the same construct
- **Content validity** → (for testing procedures) → ensuring that a test (or test battery) covers all relevant abilities needed for a particular sport (e.g., jumping, sprinting, strength)

Reliability

A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research

Terry K. Koo, PhD [a,*], Mae Y. Li, BPS [b]

Types	Definitions
Interrater reliability	It reflects the variation between 2 or more raters who measure the same group of subjects.
Test-retest reliability	It reflects the variation in measurements taken by an instrument on the same subject under the same conditions. It is generally indicative of reliability in situations when raters are not involved or rater effect is neglectable, such as self-report survey instrument.
Intrarater reliability	It reflects the variation of data measured by 1 rater across 2 or more trials.

Assessing the External Load Associated With High-Intensity Activities Recorded During Official Basketball Games

Marco Pernigoni [1,2], Davide Ferioli [3], Ramūnas Butautas [2], Antonio La Torre [1,4] and Daniele Conte [5]*

TABLE 1 | Intratester reliability of time-motion analysis variables.

	ICC (90% CI)	CV (90% CI)
Frequency (n/min)		
Sprint	0.98 (0.93–0.99)	11.94 (8.71–19.65)
HSM	0.98 (0.95–0.99)	7.29 (5.32–11.99)
Jump	1.00 (1.00–1.00)	0.00 (0.00–0.00)
Duration (s)		
Sprint	0.99 (0.97–1.00)	5.49 (4.00–9.03)
HSM	0.88 (0.67–0.96)	11.16 (8.14–18.36)
Jump	1.00 (0.99–1.00)	1.83 (1.33–3.01)

CI, confidence interval; ICC, intraclass correlation coefficient; CV, coefficient of variation; HSM, high-intensity specific movements.

Common metrics to assess reliability

- **Coefficient of variation (CV)**
  Expresses the SD as a % of the mean
  Easy to compare the amount of variation between measurements
  Value range: (any positive value)

$$CV\ (\%) = \left(\frac{Standard\ deviation}{Mean}\right) \times 100$$

- **Intraclass correlation coefficient (ICC)**
  Indicates poor (< 0.5), moderate (0.5-0.75), good (0.75-0.9) or excellent (> 0.9) reliability
  Value range: -1 to 1

$$ICC = \frac{True\ variance}{True\ variance + Error\ variance}$$

- **Pearson's product moment correlation (r)**
  High correlations may lead to the conclusion that a protocol is reliable
  Value range: -1 to 1

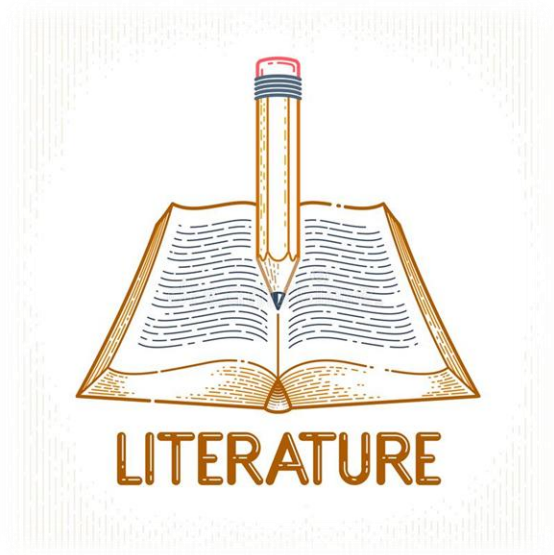
$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

Examples of validity and reliability

Over the next slides, we will go through some examples of research assessing the validity and/or reliability of different tracking/monitoring systems and testing procedures

Validity and reliability-related information

Scientific literature databases



PubMed

Clarivate
Web of Science™

Scopus

SPORTDiscus
with Full Text | Powered by EBSCOhost

Books from reliable publishing companies

REMEMBER

Books can be amazing tools to approach a new topic, but they do not go through the same rigorous peer-review process as research articles

Make sure to select books from reputable publishers

HUMAN KINETICS

Routledge
Taylor & Francis Group

Wolters Kluwer

NSCA®
NATIONAL STRENGTH AND CONDITIONING ASSOCIATION

Springer

Before we proceed

The purpose of this lecture is **NOT** to examine the specific validity and reliability of every type of tracking/monitoring system and testing procedure

But rather to provide general information on how to collect relevant information to assess these characteristics

Global Positioning Systems (GPS)

The Validity and Reliability of Global Positioning Systems in Team Sport: A Brief Review

Macfarlane T.U. Scott,[1,2] Tannath J. Scott,[2,3] and Vincent G. Kelly[1,2]

[1]The School of Human Movement and Nutrition Sciences, University of Queensland, Brisbane, Australia; [2]Performance Sciences Department, Brisbane Broncos Rugby League Club, Brisbane, Australia; and [3]Applied Sport Science and Exercise Testing Laboratory, Faculty of Science and Information Technology, University of Newcastle, Ourimbah, Australia

Practical Applications

Review of the literature suggests that all GPS units, regardless of sampling rate, are capable of athlete tracking for distance, with adequate intraunit reliability to allow multiple comparisons of a single device. However, coaches and practitioners should be aware of limitations of earlier 1Hz and 5Hz GPS units when interpreting distance during high-intensity running, velocity measures, and short linear running. To date it would seem that 10Hz GPS devices are the most valid and reliable across linear and team sport simulated running, overcoming many limitations of earlier models. Increasingly, the 15Hz GPS devices have had no additional benefit and in fact these performed worse in studies testing both 10 and 15Hz GPS devices.

Inertial Measurement Units

RESEARCH ARTICLE

Validity and reliability of an accelerometer-based player tracking device

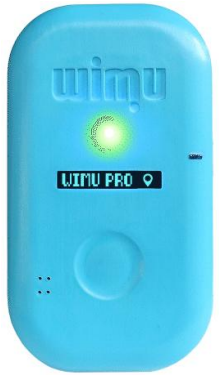
Daniel P. Nicolella[1]*, Lorena Torres-Ronda[2], Kase J. Saylor[1], Xavi Schelling[2]

The Reliability of MinimaxX Accelerometers for Measuring Physical Activity in Australian Football

Luke J. Boyd, Kevin Ball, and Robert J. Aughey

Validity and reliability of the WIMU inertial device for the assessment of the vertical jump

José Pino-Ortega[1], Javier García-Rubio[2,3] and Sergio J. Ibáñez[3]



Inertial Measurement Units

Shaker table setup: The table moves in the X (fore-aft) direction. For Y (side-side) motion, the bracket is rotated 90°. For Z (vertical) motion, the table is removed, and the bracket is attached to the motor output

Conclusions

The Catapult OptimEye S5 units showed excellent intradevice reliability. The data collected with these devices, and therefore possible decisions made using this data, will be reliable when the same device is used for each athlete over the time course of interest. However, since the interdevice reliability was shown to be highly variable (trivial to extreme), data collected on individual athletes using different devices will be of diminished reliability and utility. Therefore, it is recommended that the same device be used for the same athlete over the course of a season (or longer) in order to provide a consistent basis for comparison and to avoid interdevice variability affecting the collected data.

(from Nicolella et al., 2018)

Optojump system

Validity and Reliability of Optojump Photoelectric Cells for Estimating Vertical Jump Height

Julia F. Glatthorn,[1] Sylvain Gouge,[2] Silvio Nussbaumer,[1] Simone Stauffacher,[1] Franco M. Impellizzeri,[1] and Nicola A. Maffiuletti[1]

Table 1. Concurrent validity of force plate and Optojump cells for jump height estimation.*†

	SJ	CMJ	CMJ+
Force plate (95% CI), cm	38.9 ± 7.1 (35.6–42.2)	42.0 ± 7.1 (38.7–45.3)	48.1 ± 8.4 (44.2–52.1)
Optojump (95% CI), cm	38.0 ± 6.9 (34.7–41.2)‡	41.0 ± 7.0 (37.7–44.3)‡	46.9 ± 8.3 (43.0–50.8)‡
Systematic bias, cm	−0.9	−1.0	−1.3
Random error, cm	±1.1	±0.9	±1.1
ICC (95% CI)	0.997 (0.993–0.999)	0.998 (0.995–0.999)	0.998 (0.995–0.999)

Table 2. Test-retest reliability of Optojump cells for jump height estimation.*†

	SJ	CMJ	CMJ+
Session 1 (95% CI), cm	28.8 ± 7.4 (25.3–32.2)	31.4 ± 8.4 (27.5–35.3)	35.2 ± 10.0 (30.5–39.8)
Session 2 (95% CI), cm	29.1 ± 7.2 (25.7–32.5)	31.5 ± 8.4 (27.6–35.4)	34.8 ± 8.8 (30.7–38.9)
Systematic bias, cm	−0.32	−0.11	0.36
Random error, cm	±2.68	±2.43	±3.31
ICC (95% CI)	0.982 (0.956–0.993)	0.989 (0.973–0.996)	0.984 (0.960–0.994)
CV, %	3.1	2.2	2.8

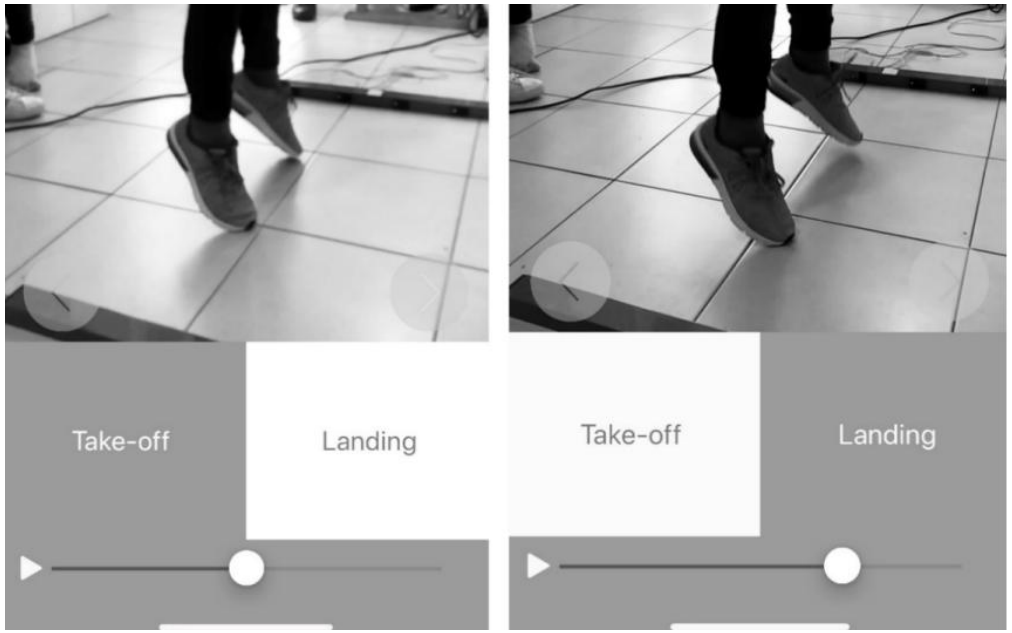
Jump height and video analysis

Validity and reliability of "*My Jump app*" to assess vertical jump performance: a meta-analytic review

Cebrail Gençoğlu[1], Süleyman Ulupınar[1]✉, Serhat Özbay[1], Murat Turan[1], Buğra Çağatay Savaş[1], Selim Asan[1] & İzzet İnce[2]

THE CONCURRENT VALIDITY AND RELIABILITY OF A LOW-COST, HIGH-SPEED CAMERA-BASED METHOD FOR MEASURING THE FLIGHT TIME OF VERTICAL JUMPS

CARLOS BALSALOBRE-FERNÁNDEZ,[1,2] CARLOS M. TEJERO-GONZÁLEZ,[1] JUAN DEL CAMPO-VECINO,[1] AND NICOLÁS BAVARESCO[2]



Smartphone / computer softwares such as *My Jump*, *Kinovea* and *Tracker* appear to be valid and reliable video analysis tools for vertical jump assessment

Heart rate variability

Validity and reliability of different smartphones applications to measure HRV during short and ultra-short measurements in elite athletes

M. Moya-Ramon [a], M Mateo-March [a,b,*], I. Peña-González [a], M. Zabala [c], A. Javaloyes [a]

Results: Compared to an electrocardiogram, Elite HRV and Welltory showed no differences neither in supine nor in seated positions ($p > 0.05$) and they showed very strong to almost perfect correlation levels ($r = 0.77$ to 0.94). Furthermore, no differences were found between short (5 min) and ultra-short (1 min) length measurements. Intraclass correlation coefficient showed good to excellent reliability and the standard error of measurement remained lower than 6%.

Conclusion: Both smartphone applications can be implemented to monitor HRV using short- and ultra-short length measurements in elite endurance athletes.

Heart rate variability

Pay close attention when using smarphone apps / smart watches / etc.

Which specific HRV indicator is being used?

Common HRV metrics: Time domain

Name	Domain	Description	Comments
SDNN (ms)	Time	Standard deviation of normal R-R (NN) intervals	Represents all cyclic variation over a reading and has limited statistical use since the measured variability increases with the recording duration
SDANN (ms)	Time	Standard deviation of average NN intervals for every 5-min cycle of a measurement	Used in longer HRV measures to estimate the variability due to cycles longer than 5 min
RMSSD (ms)	Time	Root mean square of successive differences in R-R intervals	Estimates the short-term, parasympathetic component of variability with useful statistical properties; one of the most commonly used measures in sport applications
pNN50 index (%)	Time	Percentage of consecutive NN intervals differing by more than 50 ms	Estimates the parasympathetic component of variability and correlates closely with RMSSD, though with less useful statistical utility

Common HRV metrics: Frequency domain

VLFP (ms^2)	Frequency	Absolute power of very low-frequency band (0.0033-0.04 Hz)	The physiological underpinnings are not as clearly defined as other frequency measures; questionable utility for variability measurements under 5 min
LFP (ms^2)	Frequency	Absolute power of low-frequency band (0.04-0.15 Hz)	Considered a measure of both sympathetic and parasympathetic modulation of HR and can be expressed in normalized units
HFP (ms^2)	Frequency	Absolute power of high-frequency band (0.15-0.4 Hz)	Considered a measure of parasympathetic modulation of HR (the component of variability linked to respiration) and can be expressed in normalized units
LF/HF	Frequency	Ratio of absolute power of low-frequency band to high-frequency band	Considered a measure of autonomic balance, where a high LF/HF ratio indicates sympathetic predominance
Total power (ms^2)	Frequency	The total HR power between 0.00066 and 0.34 Hz	Measures the variance of all NN intervals and is highly subject to body position and breathing rate

Custom-made indexes

Subjective assessment

Convergent Evidence for Construct Validity of a 7-Point Likert Scale of Lower Limb Muscle Soreness

Franco M. Impellizzeri, MSc† and Nicola A. Maffiuletti, PhD**

TABLE 1. Likert Scale of Muscle Soreness for Lower Limb

Value	Description
0	A complete absence of soreness
1	A light pain felt only when touched/a vague ache
2	A moderate pain felt only when touched/a slight persistent pain
3	A light pain when walking up or down stairs
4	A light pain when walking on a flat surface/painful
5	A moderate pain, stiffness or weakness when walking/very painful
6	A severe pain that limits my ability to move

Interventions: 4-week preseason training camp, which included high-intensity plyometric training sessions.

Main Outcome Measurements: Players self-reported the perceived muscle soreness of the lower limbs using the VAS (criterion measure) and the 7-point Likert scale of muscle soreness.

Results: Significant individual correlations were found between the 2 muscle soreness scales (mean r = 0.80 ± 0.07; range, 0.65 to 0.94). The correlation using the pooled data was 0.81. No significant muscle soreness scale × time interaction was found for standardized measures of muscle soreness (P = 0.98). The main factor for time (24, 48, 72, and 96 hours after the first plyometric training session) was significant (P = 0.0001). Effect sizes for the changes in the Likert and VAS absolute scores during the first 96 hours were similar (partial η^2 = 0.13).

Conclusions: The results of this study provide further convergent evidence for the construct validity of the 7-point Likert scale of muscle soreness. The 2 scales showed similar sensitivity to muscle soreness caused by eccentric contractions during the first 96 hours after plyometric exercises.

Subjective assessment

ORIGINAL RESEARCH ARTICLE

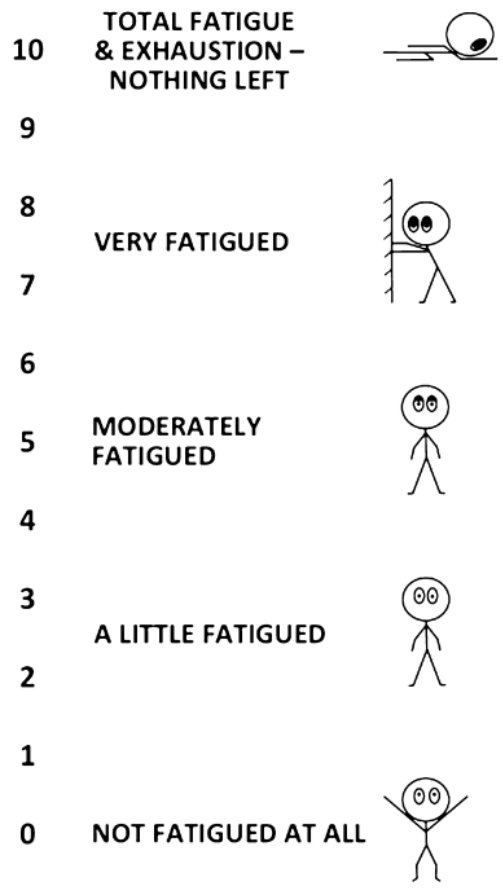
Development and Validity of the Rating-of-Fatigue Scale

D. Micklewright[1] · A. St Clair Gibson[2] · V. Gladwell[1] · A. Al Salman[3]

Key Points

A new method of measuring perceived fatigue named the rating-of-fatigue (ROF) scale has been developed.

The ROF scale was found to have good face validity and high levels of convergent validity during ramped cycling to exhaustion exercise, resting recovery and daily living activities. The ROF scale was also found to discriminate between perceived exertion during recovery from exercise.



Testing procedures

Test–retest reliability of the 30–15 Intermittent Fitness Test: A systematic review

Jozo Grgic [a,*], Bruno Lazinica [b], Zeljko Pedisic [a]

5. Conclusion

From the results of this systematic review, it can be concluded that the 30–15 IFT has excellent test–retest reliability for both maximal velocity and peak heart rate. Therefore, this test may be used as a reliable measure of fitness in research and sports practice. In particular, the 30–15 IFT can be used as a reliable test for monitoring athletes' performance and for determining the efficacy of a given training program.

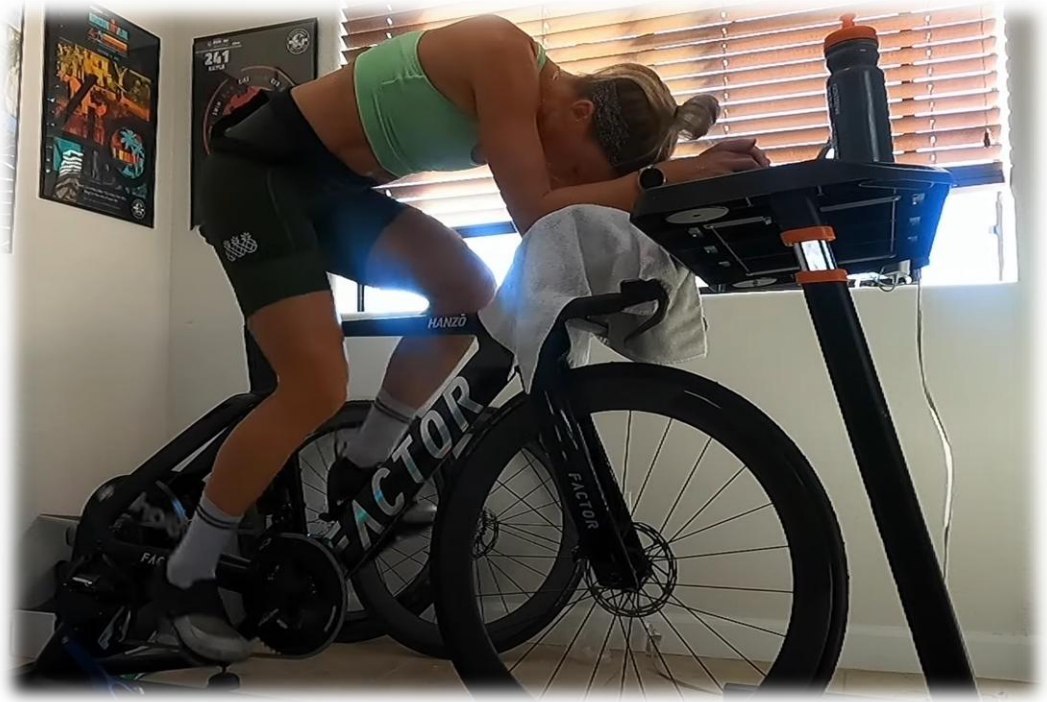
Testing procedures

PHYSIOLOGY AND NUTRITION

What is known about the FTP^{20} test related to cycling? A scoping review

Jon Mackey and Katy Horner

The FTP^{20} test was reported to be a reliable and repeatable test in all studies; by McGrath et al. (2019) (ICC = 0.98, R^2 = 0.96) in a highly trained cohort of cyclists and triathletes (mean (SD) VO_2peak; male (n = 12): 66.3 (5.5) and female (n = 7) 59.3 (6.9) $ml \cdot kg^{-1} \cdot min^{-1}$), MacInnis et al. (2019) (ICC = 0.98, R^2 = 0.95) in 8 well-trained male cyclists (VO_2peak; 68.2 (3.4) $ml \cdot kg^{-1} \cdot min^{-1}$), Borszcz et al. (2020) (CV = 2.9%, ICC = 0.97) in a cohort of 25 trained male cyclists (VO_2max 57.8 (7.3) $ml \cdot kg^{-1} \cdot min^{-1}$), and Lillo-Beviá et al. (2019) (CV = 1.2%, ICC = 0.98) in 11 trained male cyclists and triathletes (VO_2max 59.7 (3.0) $ml \cdot kg^{-1} \cdot min^{-1}$).



(once again)

A brief note:
Factors affecting validity and reliability

It’s not just about the tools

Ensuring that the tools we use for tracking, monitoring and testing are valid and reliable is essential to obtain meaningful data

however

As coaches / sport scientists / researchers, we must also ensure that we **use our monitoring / testing tools correctly**

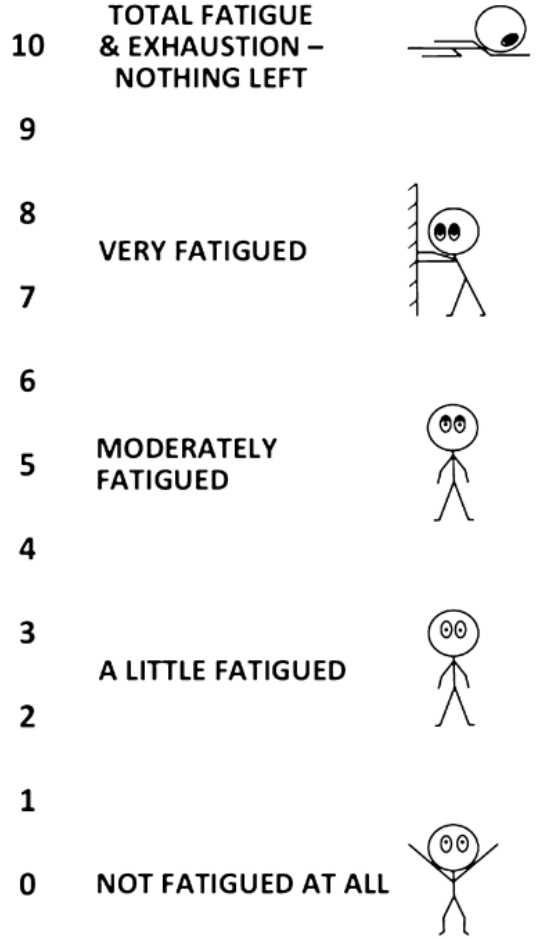
The importance of instructions

Testers / raters / observers should be well-trained to ensure that:

- Their scores correlate closely with experienced and reliable personnel
- They can **explain and administer the tests / scales as consistently as possible**
  (e.g., instructions and encouragement must be consistent!)

Instructions: Subjective monitoring

Rating-of-Fatigue scale



Understanding the Scale:

ROF ranges from 0 (“not fatigued at all”) to 10 (“total fatigue and exhaustion – nothing left”) with five descriptors and diagrams for guidance

How to Respond:

- Inspect the scale carefully, and provide a single, honest numerical response. Avoid ranges (such as “3-4”)
- Check verbal anchors first, and (only afterwards) provide a numerical value

Providing specific examples:

0 = Fully rested, like after a good night’s sleep
10 = Completely exhausted, like after sprinting to physical failure

Testing instructions: The example of CMJ

COUNTERMOVEMENT JUMP RELIABILITY PERFORMED WITH AND WITHOUT AN ARM SWING IN NCAA DIVISION 1 INTERCOLLEGIATE BASKETBALL PLAYERS

AARON D. HEISHMAN,[1,2] BRYCE D. DAUB,[2] RYAN M. MILLER,[1] EDUARDO D.S. FREITAS,[1] BRETT A. FRANTZ,[2] AND MICHAEL G. BEMBEN[1]

Article

Differences in Vertical Jump Force-Time Characteristics between Stronger and Weaker Adolescent Basketball Players

Christopher Thomas [1,*], Irene Kyriakidou [2], Thomas Dos'Santos [1] and Paul A. Jones [1]

- Distribute weight equally between feet
- Use maximal effort ("jump as high and as fast as possible")
- Avoid "stutter-steps"
- Avoid leg kicking/tucking
- Land on the same spot as take-off

Testing: Athlete familiarization

Instructions:

- Test purpose
- What the test consists of + how the scoring works
- Recommended warm-up
- Number of practice attempts
- Number of trials
- Criteria for valid attempt
- Recommendations for maximizing performance

A familiarization session is highly recommended

Testing: Environmental conditions

External factors can also influence the reliability of our testing procedures:

- **Temperature** → hot + humid, or cold weather may impair performance
- **Altitude** → until 2700 m, VO_{2max} declines by 5% every 900 m, more sharply over 2700 m

Solutions → e.g., documenting the environmental conditions

Summary: Validity and reliability in sports

1. A great variety of tracking, monitoring and testing tools are available to support the coaching process
2. The validity and reliability of these tools must be adequate, in order to obtain meaningful data
3. External factors can affect validity and reliability, which should be minimized whenever possible

Questions?