

Sports science experiments measure people, and people vary from trial to trial and from one another. The marks go to a standardised protocol, enough trials and participants, the right statistic, and an honest account of what the data can and cannot show.

Measuring well

Validity – does the test measure what it claims to? A sit-and-reach test measures hamstring and lower-back flexibility, not whole-body flexibility; a multistage fitness test estimates VO₂max, it does not measure it

Reliability – does the test give the same result when repeated under the same conditions? Check it with a test–retest: the same participants, the same protocol, a few days apart

Accuracy and precision – accuracy is closeness to the true value (a calibrated heart-rate monitor); precision is how close repeated readings are to each other (a small standard deviation)

Standardised protocol – the same warm-up, time of day, surface, footwear, instructions and encouragement for every participant and every trial. Give a familiarisation trial first so learning does not look like improvement

Name the instrument and its resolution: a stopwatch to 0.01 s (but reaction time adds about 0.2 s), light gates, a heart-rate monitor in beats per minute, a sit-and-reach box to 0.5 cm, a stadiometer to 0.1 cm, scales to 0.1 kg.

Statistics for an investigation

Mean and standard deviation – the SD describes the spread of the data about the mean; about 68% of normally distributed values lie within 1 SD of the mean

Coefficient of variation – CV = SD ÷ mean × 100%. It compares the spread of data sets with different means – the variability of sprint times against jump heights

t-test t – is the difference between two means significant? Paired for the same people tested twice (before and after training), unpaired for two different groups. p < 0.05 means a difference this large would arise by chance less than 5% of the time

Correlation – Pearson’s r (or Spearman’s for ranks) from –1 to +1 measures how closely two variables are linearly related; r² is the fraction of the variation in one explained by the other. Correlation does not show cause

Error bars of ± 1 SD on a bar chart: bars that overlap widely suggest the difference may not be significant; run the test to decide.

Worked calculation: reaction time from a ruler drop

A partner holds a metre rule with its zero between the participant’s open finger and thumb, then drops it without warning; the participant catches it as fast as possible. The distance fallen d was read at the top of the thumb. Five trials after one practice trial, using the dominant hand.

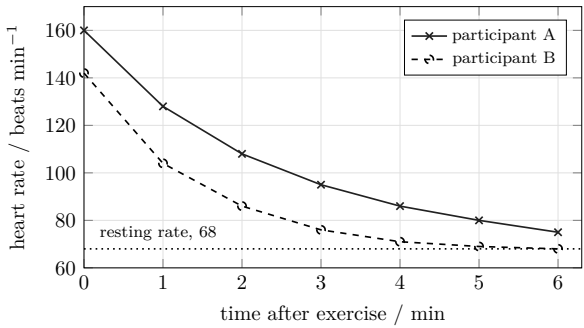
trial	1	2	3	4	5	mean
distance fallen / cm	18.5	21.0	19.0	17.5	20.0	19.2

- The rule falls freely from rest, so $d = \frac{1}{2}gt^2$, which gives $t = \sqrt{(2d \div g)}$.
- $t = \sqrt{(2 \times 0.192 \text{ m} \div 9.81 \text{ m s}^{-2})} = \sqrt{0.0391} = 0.198 \text{ s}$, about 0.20 s.

The partner must not signal the drop – vary the delay before releasing. Use the mean of several trials: one anticipated catch gives a time far too short, which is an anomaly to leave out.

Worked analysis: heart-rate recovery

Two participants with the same resting heart rate did the same 3-minute step test (step height, rate set by a metronome). Heart rate was read from a chest-strap monitor every minute afterwards, seated.



Points joined in order: each reading is a separate measurement in time.

Describe: B’s heart rate rose less (142 against 160 beats per minute) and returned to rest within about 5 minutes; A’s was still 7 beats per minute above resting after 6 minutes.

Explain: B is likely to be aerobically fitter. A larger stroke volume means the same cardiac output at a lower heart rate, and a faster recovery shows the oxygen debt (EPOC) is repaid sooner.

Limitations: one trial for each person; two participants cannot support a general claim; the step test suits some leg lengths better than others. Improvements: several participants in each group, a test–retest of the step test, and a t-test on the recovery times.

Ethics and safety

Informed consent in writing (from a parent for under-18s), a health screening questionnaire such as the PAR-Q before any maximal test, the right to withdraw at any time without giving a reason, and anonymous data.

Name the risk in THIS test and its control: a warm-up and cool-down for a maximal sprint; a spotter and a non-slip surface for a jump test; stop a multistage test at once if the participant feels dizzy or in pain.

What examiners keep reporting

Confusing reliability with validity: a test can give the same wrong answer every time – reliable but not valid.

A protocol without its details: ‘participants warmed up’ gains nothing; ‘a 5-minute jog at a self-selected pace, then four dynamic stretches’ does.

Too few participants or trials to support the conclusion, and no statement of that limitation.

Calling a correlation a cause, or a non-significant difference ‘a small effect’ without a statistical test.