

Middlesex University
London Sport Institute

@ProfAnthonyTurner

Worksheets and videos available here

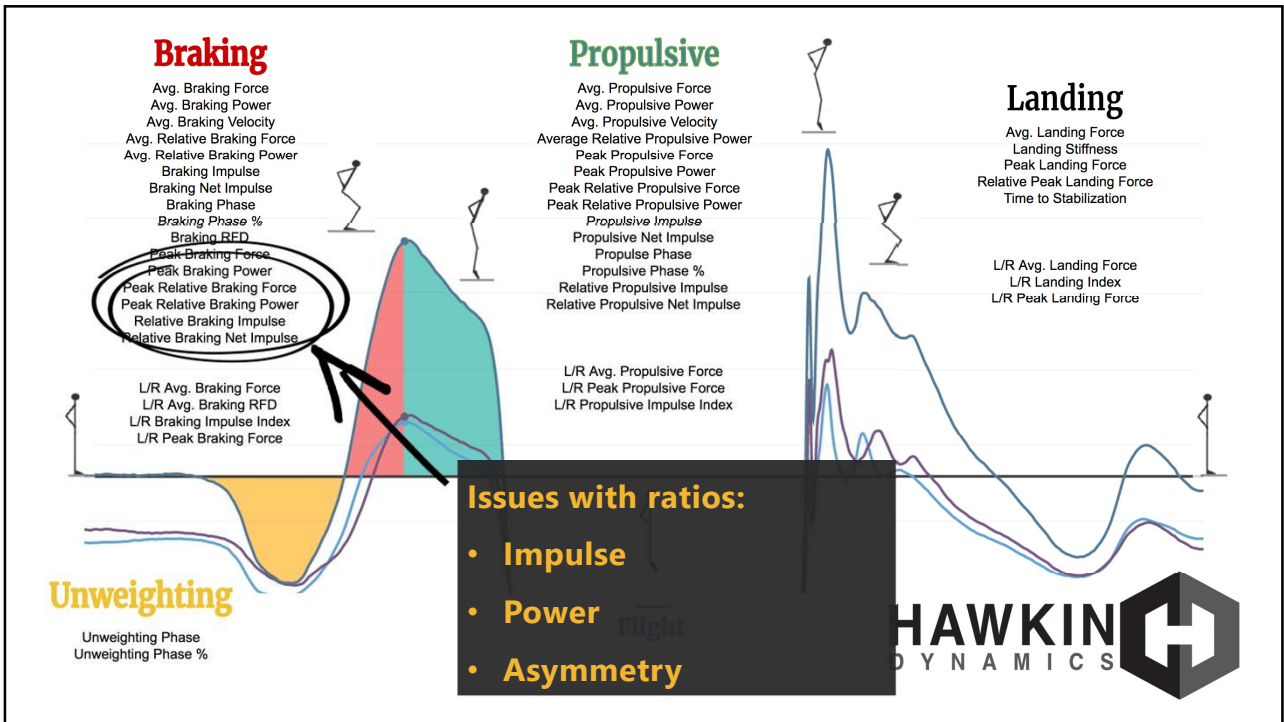
**Determining Meaningful Change in Metrics that Matter:
Data analysis when N = 1**

Prof. Anthony Turner | a.n.turner@mdx.ac.uk <https://thefitnessformula.training/workshops>

Presentation aim. How to:

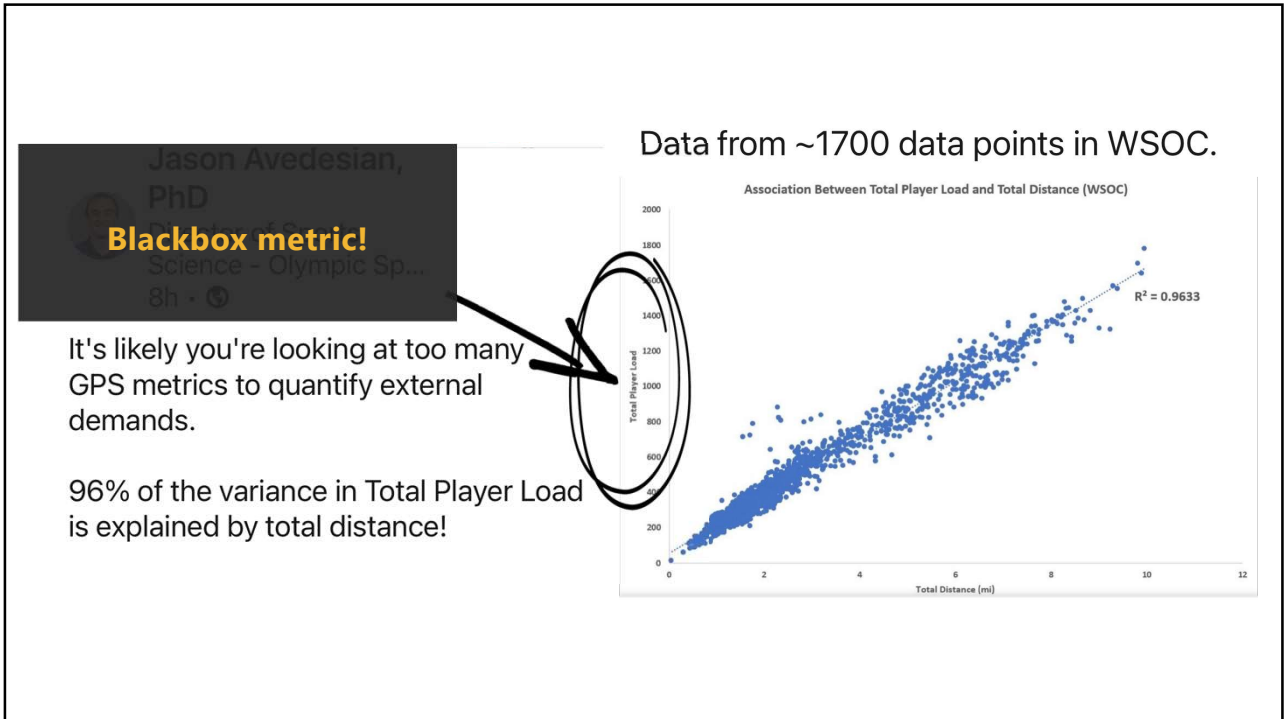
- 1 Identify metrics that inform your practice
- 2 Standardize the test (apples to apples)
- 3 Determine how much you can trust any metric
- 4 N = 1 for performance (capability)
- 5 N = 1 for Readiness (stability)
- 6 Benchmarking against teammates (TSA)

Step 1. Select the metrics



Athlete	Peak power (W)	Force at peak power (N)	Velocity at peak power (m·s ⁻¹)	How they got 4500 W
A	4500	3000	1.50	Force-led (big force, moderate velocity)
B	4500	2500	1.80	Balanced
C	4500	2250	2.00	Velocity-led
D	4500	2000	2.25	More velocity-led
E	4500	1800	2.50	Very velocity-led

ATHLETE	LEFT FORCE (N)	RIGHT FORCE (N)	ASYMMETRY %	INTERPRETATION (ASYM%)	INTERPRETATION (FORCES)
A	1000	1200	18%	High asymmetry — a concern	Both legs strong; asymmetry less concerning
B	600	620	3%	Low asymmetry — seems fine	Both legs weak; higher risk despite symmetry
C	800	960	17%	High asymmetry — a concern	Moderate force; both asymmetry and capacity matter
D	1400	1430	2%	Low asymmetry — no concern	Very strong athlete; asymmetry irrelevant
E	500	650	26%	Very high asymmetry — major concern	Both weak and imbalanced — highest risk



Step 2. Standardize the test

...And what happens when you don't

Standardise the test to:

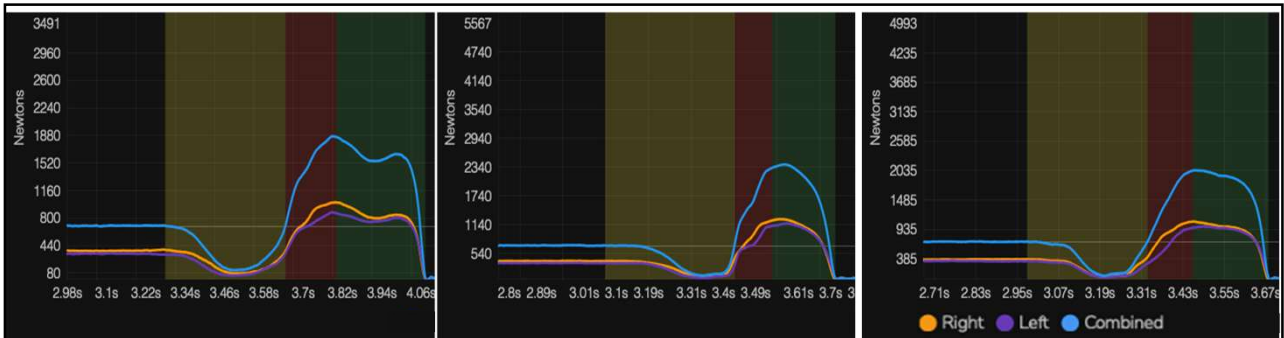
1. Isolate the muscle/s
2. Compare apples with apples

Step from the box...

Hands on hips

Walk like Lego (with arms on hips!)

Trial	Drop Height	Avg. Braking Force
FIRST TRIAL	0.17 m	2454 N
SECOND TRIAL	0.35 m	3053 N
STEP OFF AS A LEGO	0.3 m	2920 N



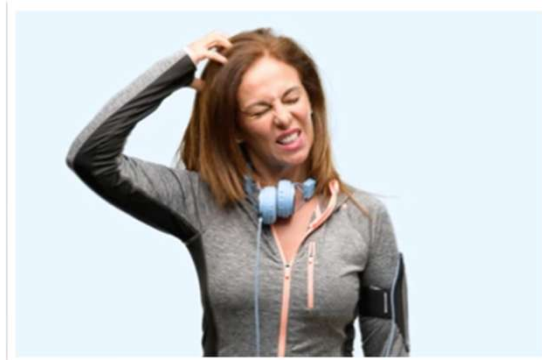
Metric	“high as you can”	“fast as you can”	“Fast and high”
Height (cm)	51	43	49
CM depth (cm)	41	25	31
Ave Braking F (N)	1477	1669	1496
Braking phase (s)	0.15	0.11	0.13
Ave Propulsive F (N)	1512	1866	1697
Propulsive phase (s)	0.27	0.18	0.22
TTT	0.79	0.64	0.7
RSImod	0.64	0.67	0.7

Step 3. Is the metric reliable
 ...How much should I trust it?

Explain this to your athlete

You bench press 3 times in a week

- In session 1 you bench 70 kg
- In session 2 you bench 72 kg
- In session 3 you bench 69 kg



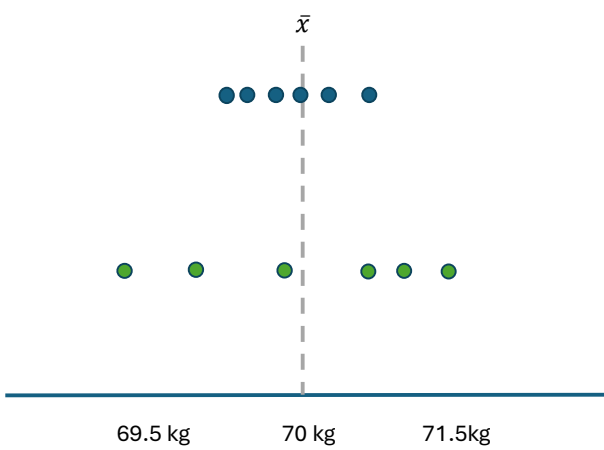
What about this...

You weigh yourself everyday for 5 days

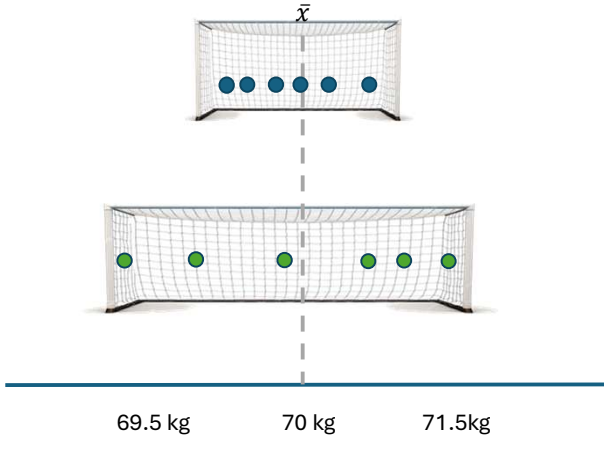
- On day 1 you weigh 70 kg
- On day 2 you weigh 70.5 kg
- On day 3 you weigh 69.9 kg
- On day 4 you weigh 70.1
- On day 5 you weigh 70.3

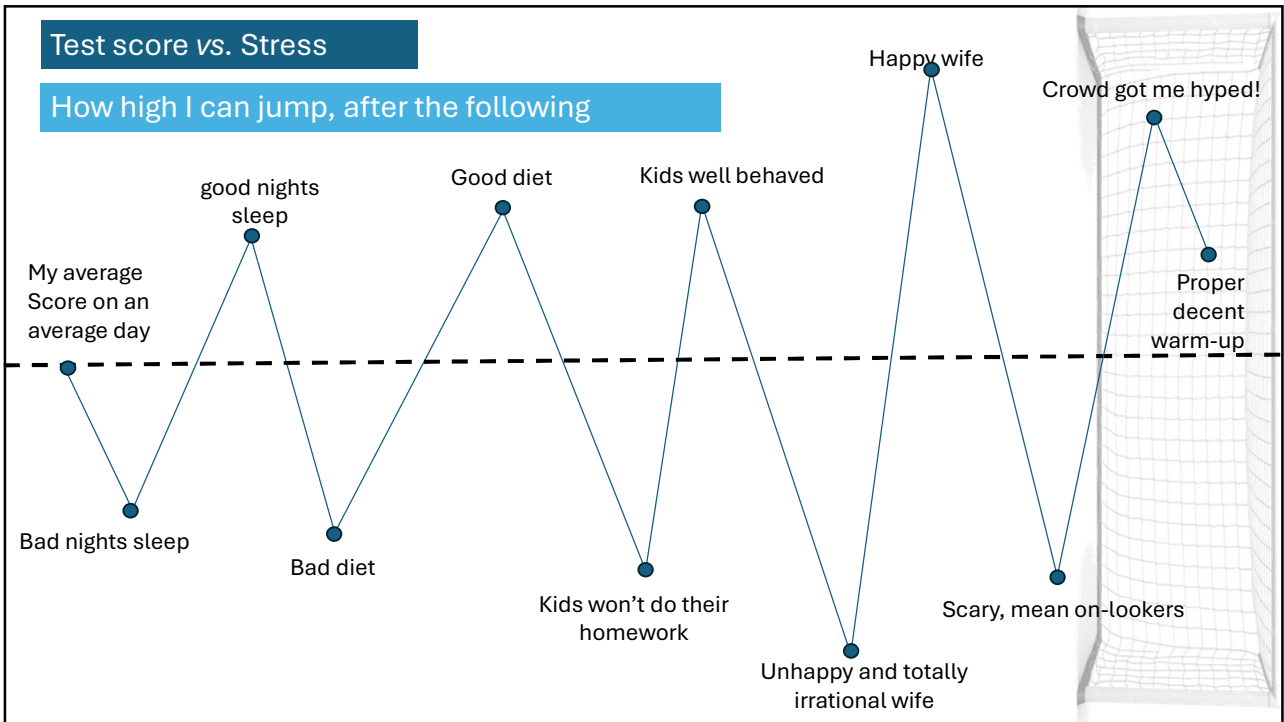
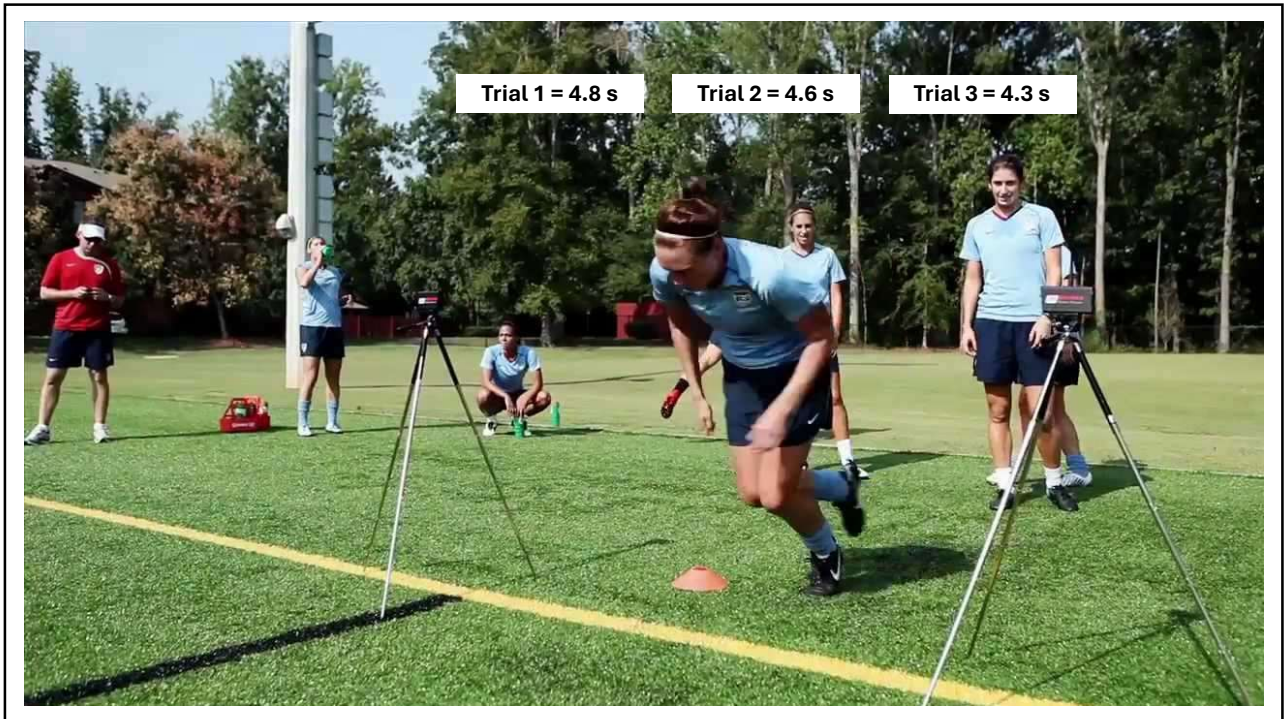


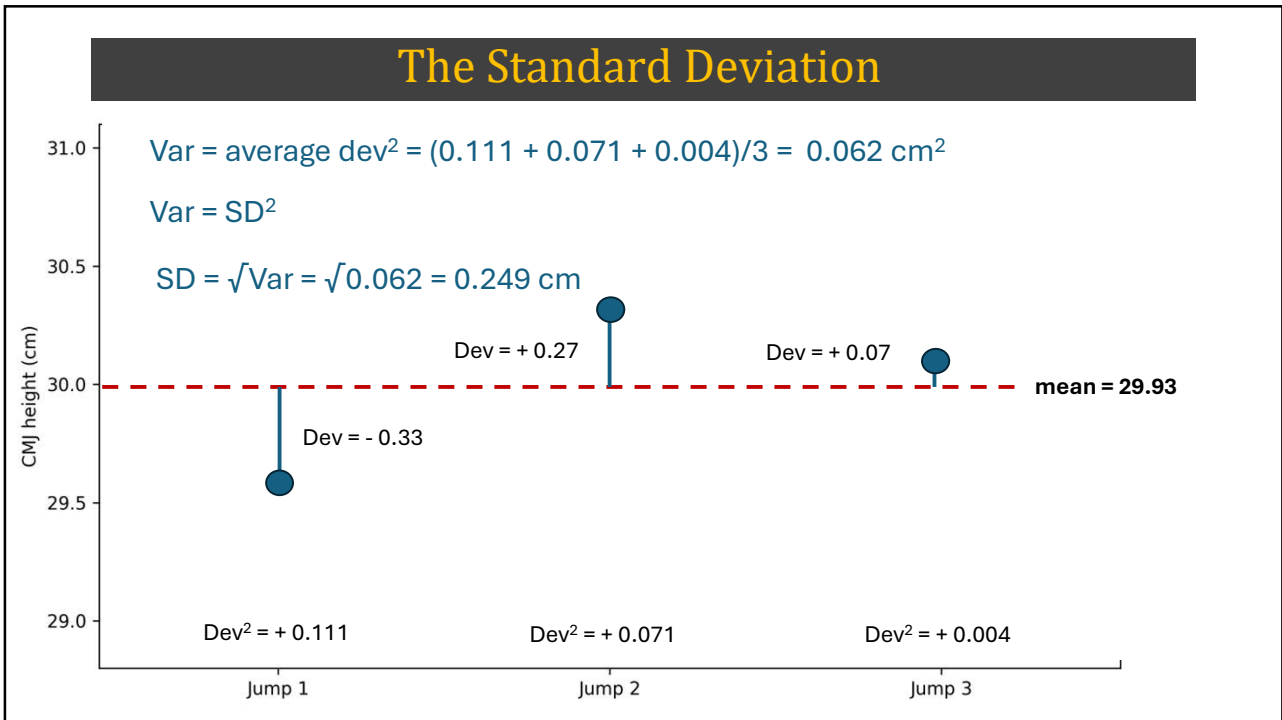
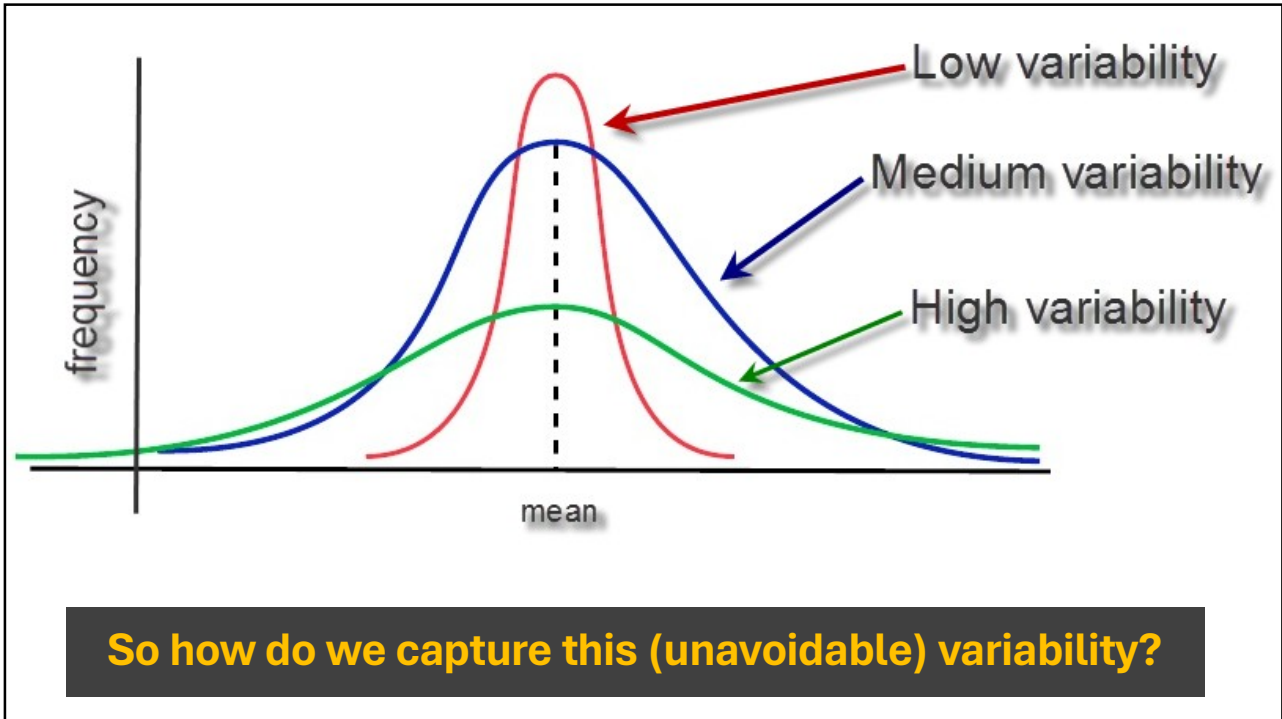
Which weighing scale would you buy?



Which weighing scale would you buy?





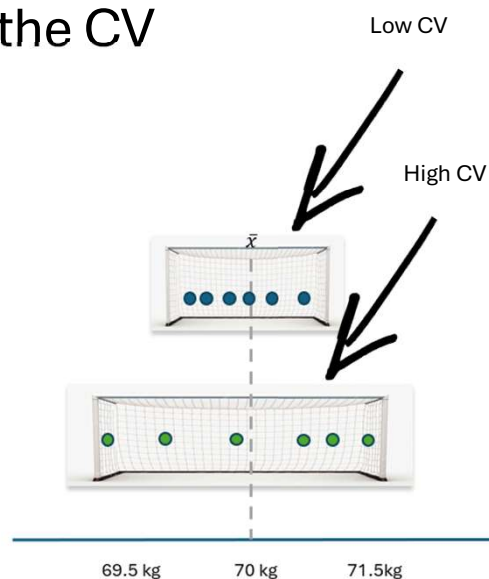


SD relative to the mean = the CV

The CV is an intuitive representation of variation

It is therefore our first line of defence

The higher the CV the more varied the data points



Original Research *Journal of Strength and Conditioning Research*

Intra- and Interday Reliability of Weightlifting Variables and Correlation to Performance During Cleans

Angela M. Sorensen,¹ Shyam Chavda,¹ Paul Comfort,² Jason Lake,³ and Anthony N. Turner¹
¹London Sports Institute, Middlesex University, London, United Kingdom; ²Human Performance Laboratory, University of Salford, Salford, United Kingdom; and ³Department of Sport and Exercise Sciences, University of Chichester, Chichester, United Kingdom

Level one: Which variables are reliable?

Level 2: Which variables are highly correlated (multicollinearity)?

Level 3. Of the correlated variables which one statistically or logically best explains the performance outcome

Results

Sixteen of the 70 variables analyzed were found to have good to excellent intra- and interday ICC (0.779–0.994 and 0.969–0.996, respectively) and CV (0.64–6.42 and 1.14–6.37, respectively) values (30,36). Using the Pearson’s correlation coefficients ($r = 0.5–1.0$ at $p < 0.005$), these 16 variables were also shown to have strong correlations ($r = 0.880–0.988$) to cleans performed at 90% 1RM. From these 16 variables, bar work variables that were used to calculate bar power variables were then excluded because they are derived from the same force and displacement data and represented duplicate data. The resulting variables were further assessed for multicollinearity, which can be seen in Table 3. This system of filtering resulted in a total of 11 variables exhibiting “good to excellent” ICC with a CV of $\leq 10\%$ for both intraday and interday reliability measures and with correlations to clean performance as reported in Table 2.

Table 2. Reliability of countermovement jump concentric force–time and |

	Trial 1 Mean ± SD	Trial 2 Mean ± SD	%CV (LCI, UCI)
CON Duration (ms)	2.80 ± 0.50	2.86 ± 0.58	4.67 (3.84, 5.49)
CON Impulse (N·s)	158.46 ± 45.23	158.77 ± 45.82	1.62 (1.03, 2.20)
CON Mean Force (N)	1273 ± 319	1265 ± 321	2.15 (1.81, 2.48)
CON Mean Power (W)	1633 ± 544	1630.7 ± 553	3.20 (2.55, 3.86)
CON Peak Force (N)	1605 ± 407	1612 ± 418	2.91 (2.44, 3.37)
CON Peak Velocity (m·s⁻¹)	2.38 ± 0.26	2.38 ± 0.28	1.44 (0.89, 1.99)
CON RFD (N·s⁻¹)	720 ± 1138	759 ± 1244	76.45 (66.86, 86.03)
CON RFD (N·m⁻¹)	14,673 ± 6521	14,308 ± 6599	7.15 (6.08, 8.21)
Jump Height (cm)	27.30 ± 6.41	27.46 ± 6.36	2.92 (2.49, 3.35)
Jump Height (Imp-Mom) (cm)	26.01 ± 6.40	26.17 ± 7.09	3.21 (2.09, 4.33)
Jump Height (Imp-Mom) (cm)	25.96 ± 6.39	26.13 ± 7.09	3.20 (2.09, 4.33)
Lower Limb Stiffness (N·m ⁻¹)	4971 ± 3081	6181 ± 10,643	10.21 (7.59, 12.82)
Peak Power (W)	2979 ± 972	2955 ± 982	2.48 (1.88, 3.08)
RSI-modified (m·s ⁻¹)	0.37 ± 0.11	0.37 ± 0.12	5.81 (4.87, 6.75)

Pick one

Are you calculating jump momentum?

WTF!

Same thing

Only if you take height and TTT

Variability helps us define thresholds for meaningful change in follow-up tests.

But when we compare two test scores, we have two sets of variability to account for

We use propagation of variance to add the variability from both tests.

And we compare the **signal** relative to the **noise**?

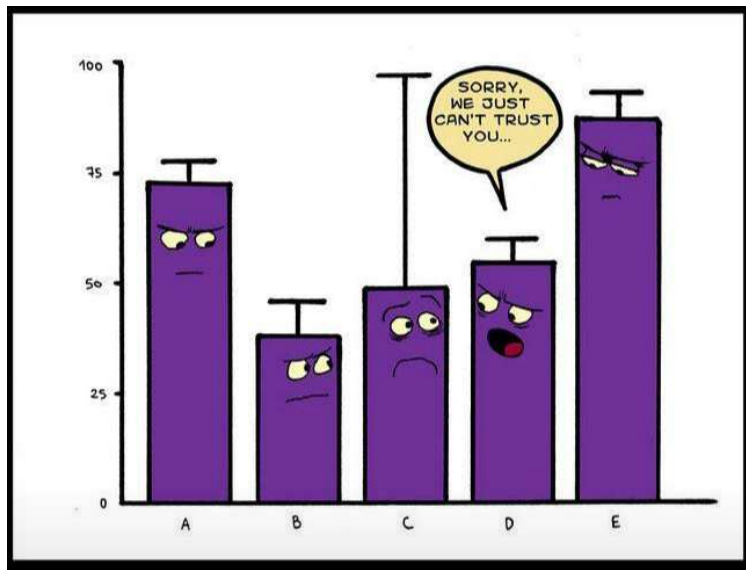
For example, consider group-based probability (NHST) statistics

$\frac{\text{signal}}{\text{noise}} = \frac{\text{difference between group means}}{\text{variability of groups}}$
 $= \frac{\bar{X}_T - \bar{X}_C}{SE(\bar{X}_T - \bar{X}_C)}$
 $= \text{t-value}$

$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$t = (45 - 41) / \text{SQRT}(5^2/16 + 5^2/16)$
 $t = 4 / \text{SQRT}(3.125)$
 $t = 2.26, p = 0.031$

Probability of getting this value, assuming no difference between groups is 3%



How to reduce the noise (**SD**)

The tester

- Expert
- Strict
- Coaching cues

Post CMJ – Pre CMJ Variability

The environment

- Temperature
- Audience
- competition
- Music



The athlete:

- Homogenous group
- Technique
- Motivation
- Biological variability

The equipment

- Recording frequency
- Calibration
- Unobtrusive

Step 4. Individual athlete analysis
...determining meaningful change
in performance

Combining uncertainty across two tests: Propagation of variance

The SD of the difference (i.e., expected noise in the change score) is therefore:

$$SD_{\Delta} = \sqrt{SD_1^2 + SD_2^2}$$

How much change would you expect to see purely based on variability

Once SD_{Δ} is known, the change can be expressed relative to its expected noise (**signal-to-noise ratio**):

$$\text{Noise scaled}_{\Delta} = \frac{\text{Perf}_1 - \text{Perf}_2}{SD_{\Delta}}$$

- Stable: $| \text{Perf}_{\Delta} | < 1$
- Likely change: $1 \leq | \text{Perf}_{\Delta} | < 2$
- Meaningful change: $| \text{Perf}_{\Delta} | \geq 2$

- CMJ height₁: 37 cm | CMJ height₂: 40 cm | $SD_1 = 0.8$ cm | $SD_2 = 0.7$ cm
- Performance change: $\Delta = 3$ cm
- Expected noise (Propagation of Variance):

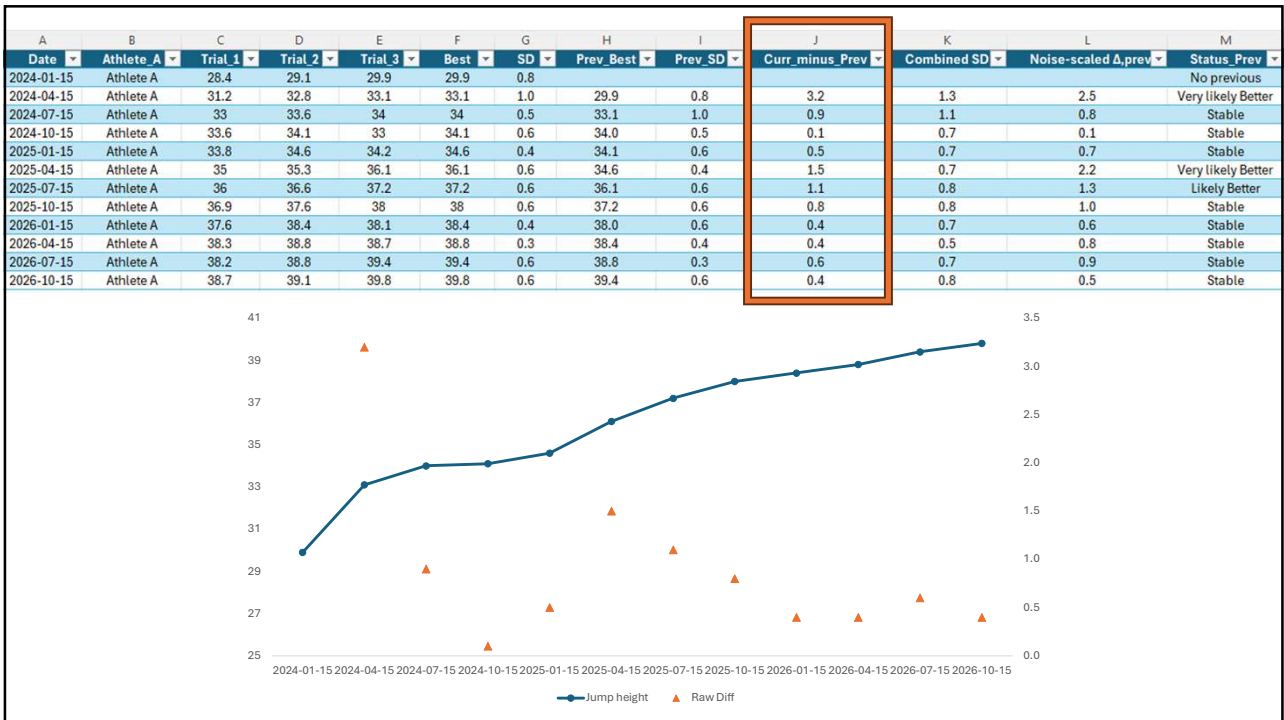
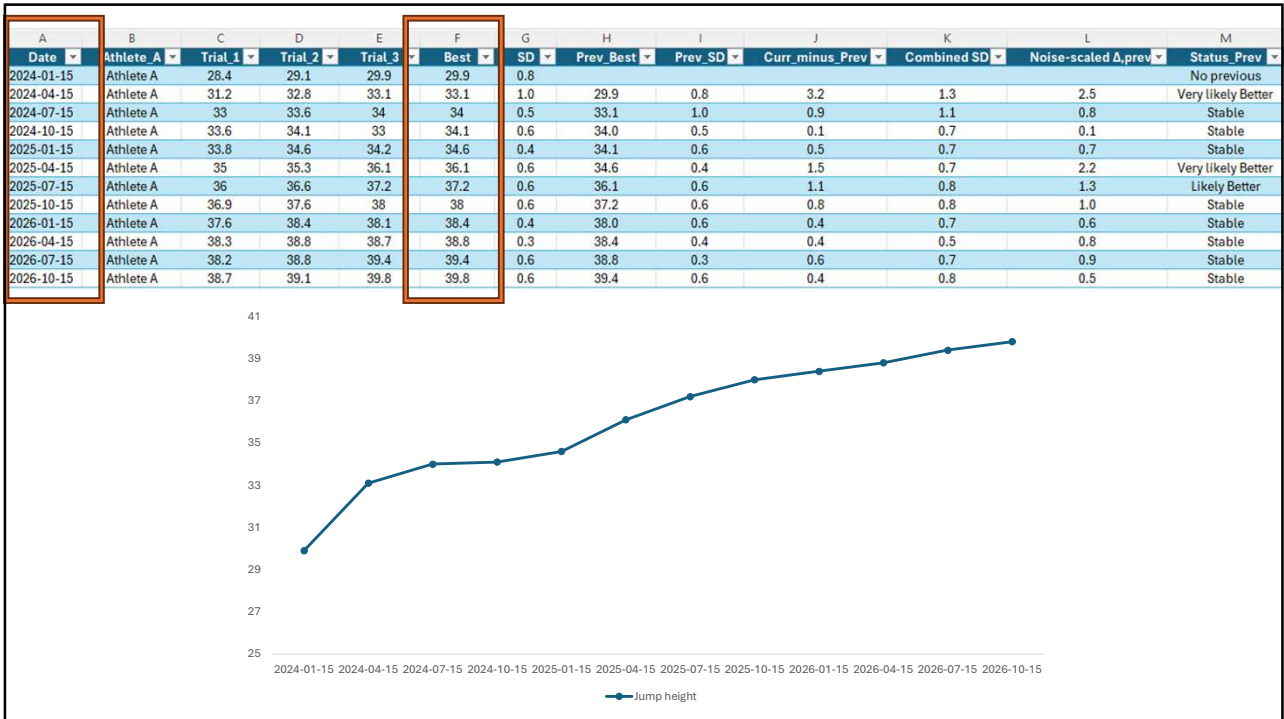
$$SD_{\Delta} = \sqrt{0.8^2 + 0.7^2} = 1.06 \text{ cm}$$

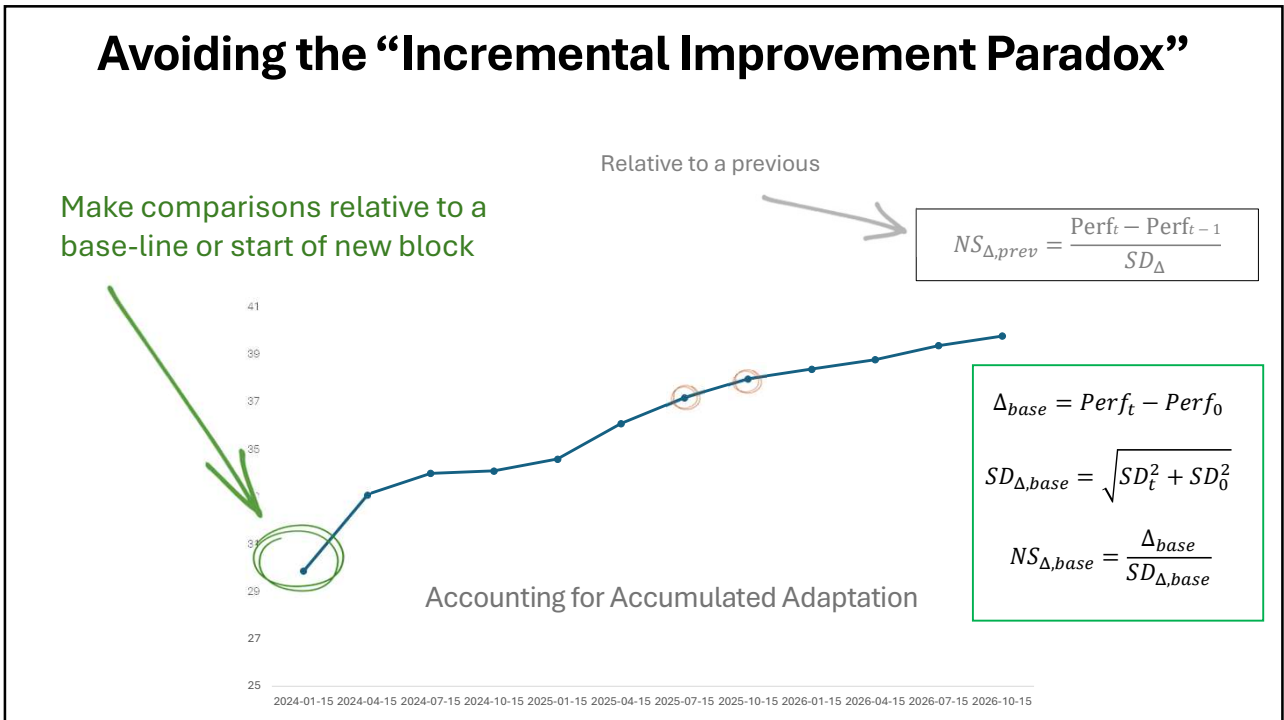
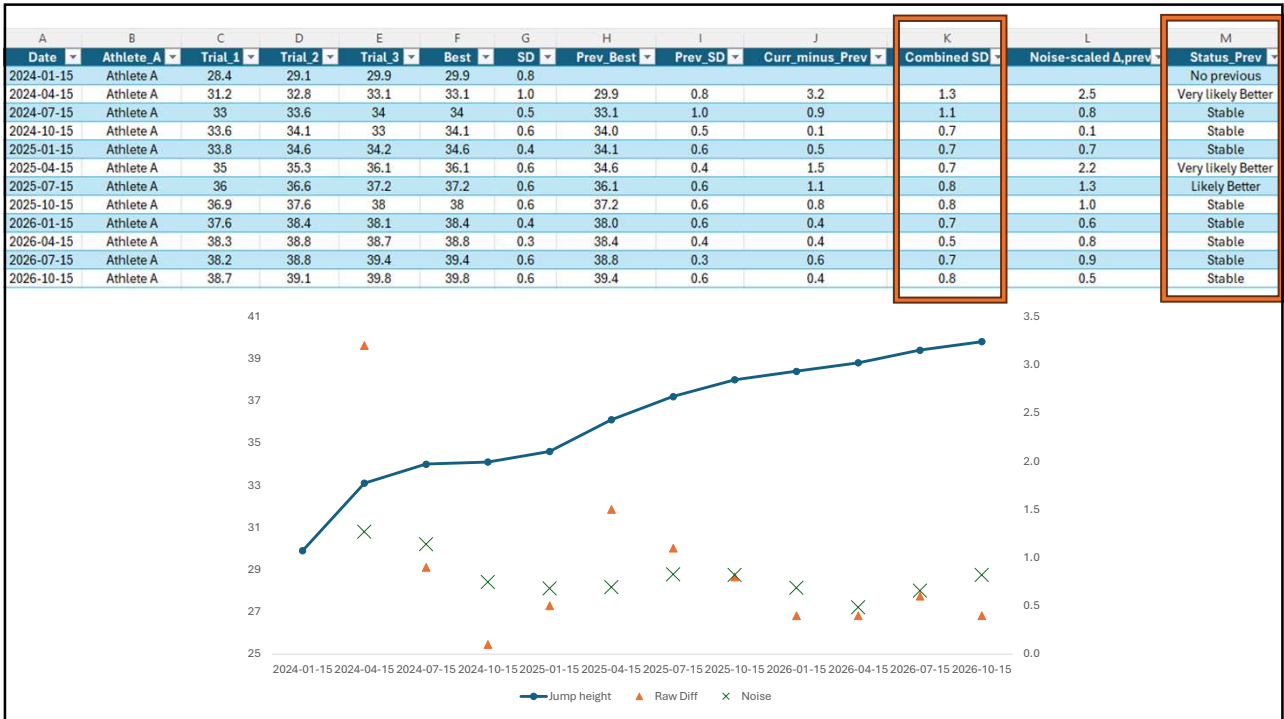
- Noise-scaled change:

$$\frac{3}{1.06} = 2.83$$

The change in performance is almost 3 times larger than what we would expect from variability alone

- Meaningful change (≥ 2)





Step 5. Individual athlete analysis

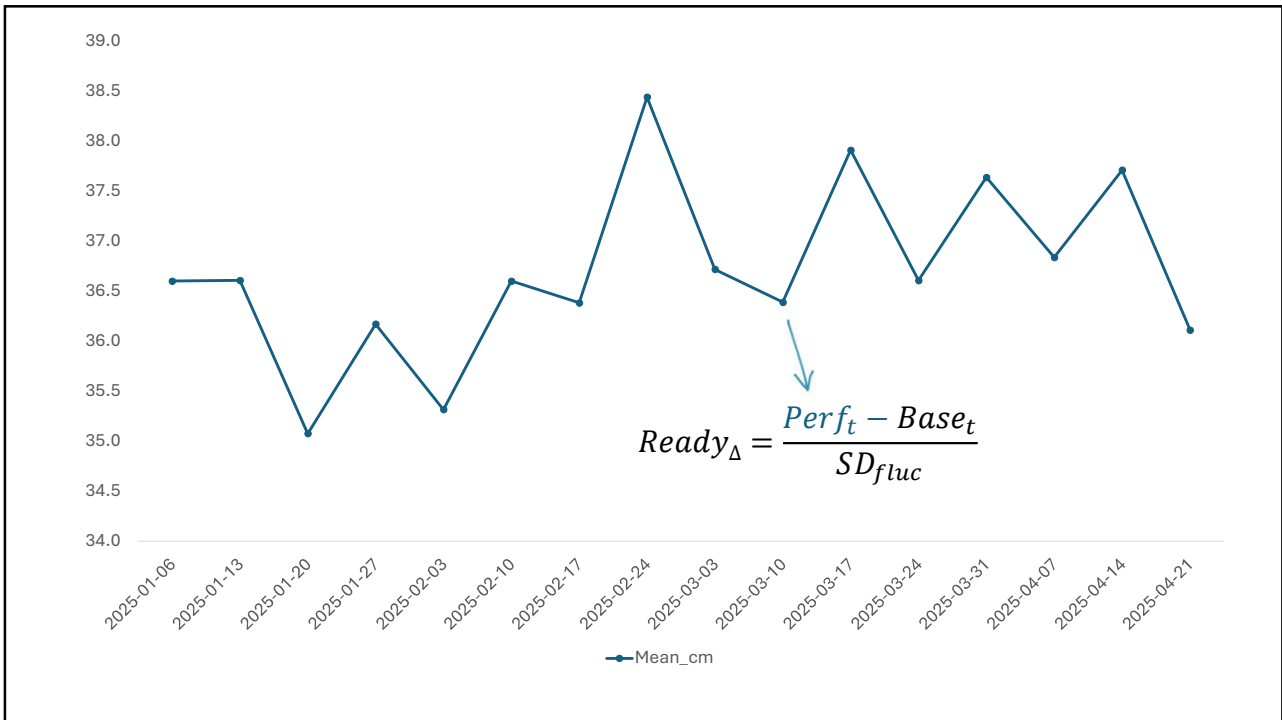
...determining meaningful change in readiness monitoring

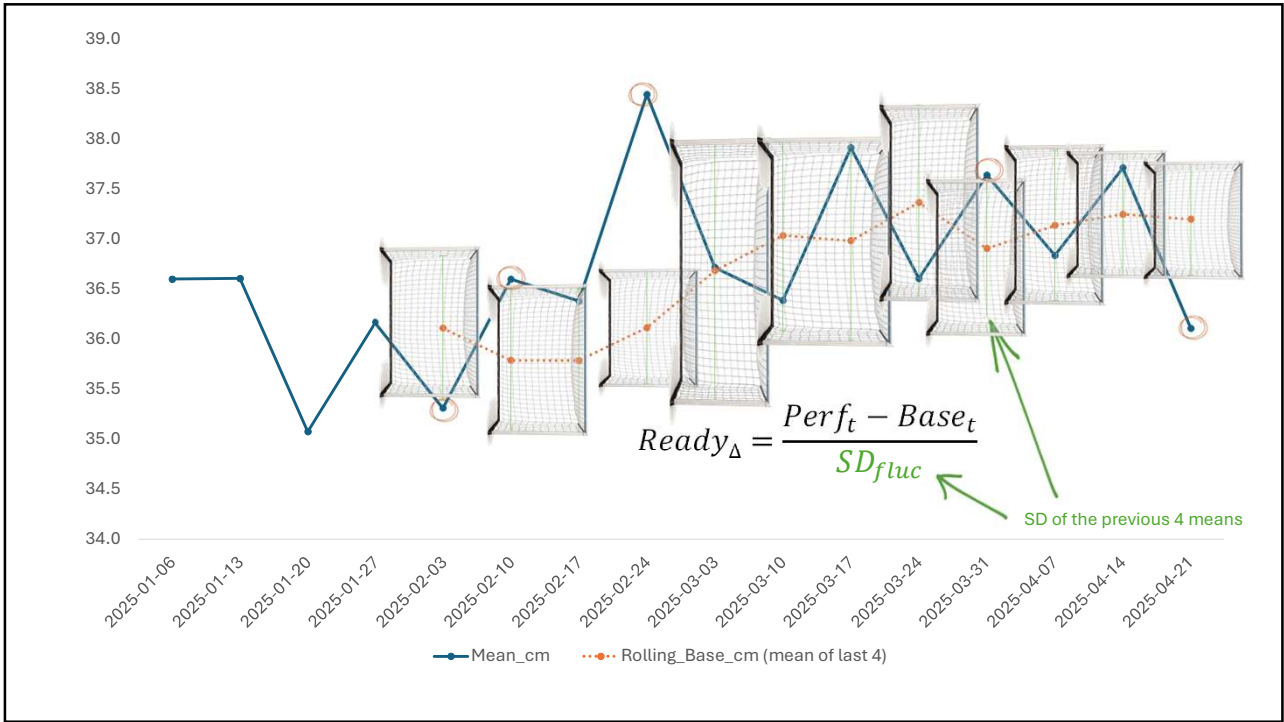
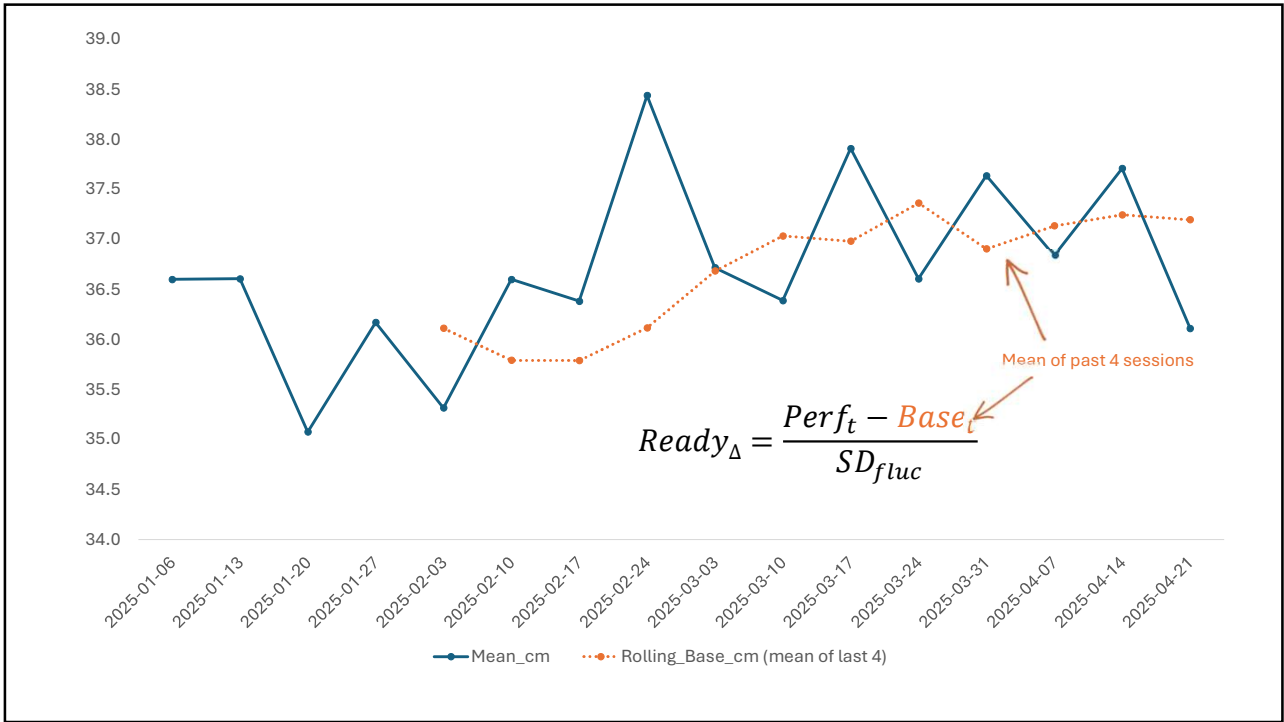
Comparison Dimension	Performance Testing	Readiness Monitoring
Focus	Capability change	State stability
Testing Frequency	Infrequent assessments	Frequent assessments
Reference Point Mean of past 4 sessions	Baseline often fixed (start-of-block/season)	Baseline dynamic (rolling window)
Goal	Detect adaptation	Detect fluctuations
Interpretation	Accumulated improvement	Transient fluctuation

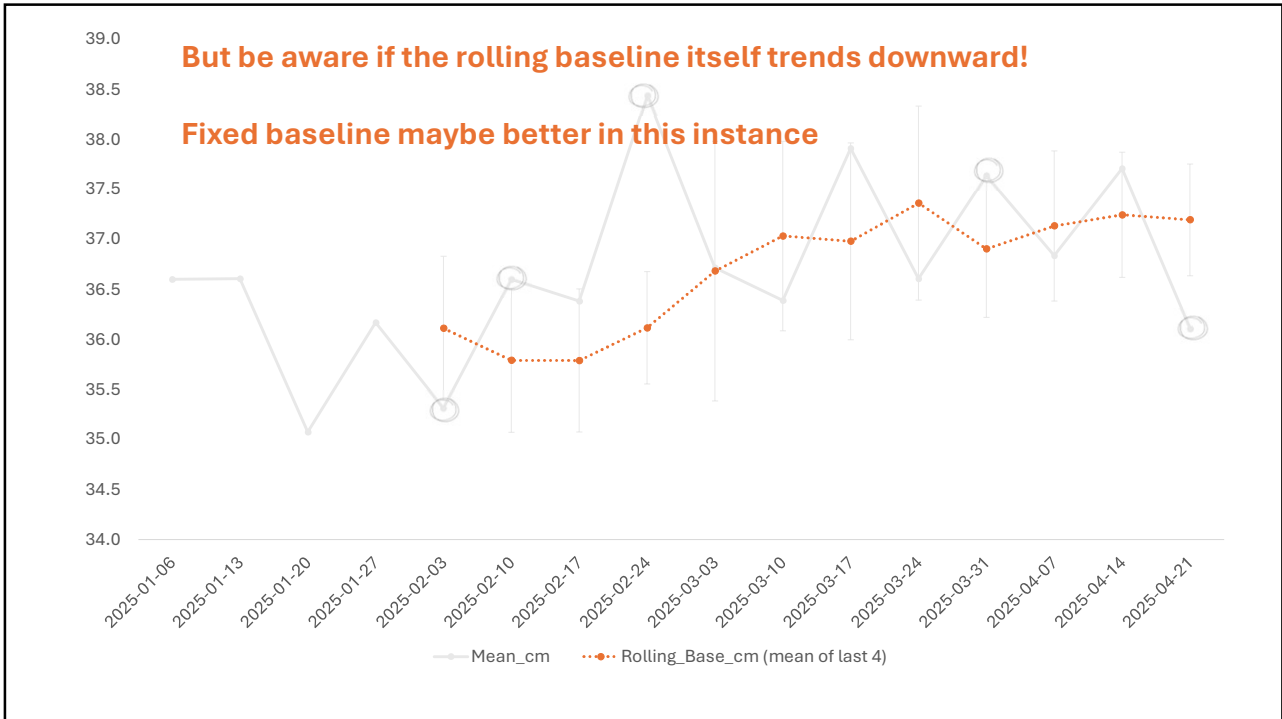
Mean

SD of the previous 4 means

	A	B	C	D	E	F	G	H	I	J	K
1	Date	Trial 1	Trial 2	Trial 3	Mean	SD	Roll Base mean last 4	Roll Fluc (SD of mean of last 4)	Curr-Roll Base	Readiness (Δ /fluc)	Status
2	2025-01-06	37.35	36.54	35.9	36.6	0.73					Insufficient history
3	2025-01-13	36.12	36.49	37.2	36.6	0.55					Insufficient history
4	2025-01-20	34.56	35.41	35.25	35.1	0.45					Insufficient history
5	2025-01-27	36.4	35.6	36.5	36.2	0.49					Insufficient history
6	2025-02-03	36.44	34.95	34.89	35.4	0.88	36.1	0.7	-0.7	-0.9	Normal
7	2025-02-10	36.03	37.04	36.72	36.6	0.52	35.8	0.7	0.8	1.1	Likely Increase
8	2025-02-17	37.42	36.02	35.7	36.4	0.91	35.8	0.7	0.6	0.8	Normal
9	2025-02-24	38.94	38.68	37.69	38.4	0.66	36.1	0.5	2.3	4.5	Meaningful Increase
10	2025-03-03	36.28	36.33	37.53	36.7	0.71	36.7	1.3	0.0	0.0	Normal
11	2025-03-10	35.9	36.9	36.36	36.4	0.50	37.0	0.9	-0.6	-0.7	Normal
12	2025-03-17	37.37	38.3	38.05	37.9	0.48	37.0	1.0	0.9	0.9	Normal
13	2025-03-24	36.8	37.1	35.91	36.6	0.62	37.4	1.0	-0.8	-0.8	Normal
14	2025-03-31	37.69	37.92	37.29	37.6	0.32	36.9	0.7	0.7	1.1	Likely Increase
15	2025-04-07	36.85	37.24	36.41	36.8	0.42	37.1	0.7	-0.3	-0.4	Normal
16	2025-04-14	38.2	37.55	37.37	37.7	0.44	37.2	0.6	0.5	0.7	Normal
17	2025-04-21	35.91	36.61	35.8	36.1	0.44	37.2	0.6	-1.1	-1.9	Likely Drop







**Step 6. Comparison with teammates
...to set realistic targets**

z-scores and the TSA



Total Score of Athleticism: Holistic Athlete Profiling to Enhance Decision-Making

Anthony N. Turner, PhD,¹ Ben Jones, PhD,¹ Perry Stewart, MSc,¹ Chris Bishop, MSc,¹ Nimal Pinar, PhD,¹ Shyam Chavda, MSc,² and Paul Reed, PhD³
¹London Sports Institute, Middlesex University, Allianz Park, London, United Kingdom; ²Carnegie Applied Rugby Research (CARR) Centre, Institute for Sport, Physical Activity and Leisure, Leeds Beckett University, Leeds, United Kingdom; ³National Performance and Research Team, Arsenal Football Club, United Kingdom; and ⁴Aspetar Orthopaedic and Sports Medicine Hospital, Doha, Qatar

ABSTRACT

Often, the various coaching staff, sport scientists, and medical practitioners of a sports club require a single, holistic indication of an athlete's abilities. Currently, there is no consensus on how this is best defined, and thus, a total score of athleticism (TSA) may provide one such method. The TSA is derived from the average of z-scores on 7 factors in the case of small samples from a sport-specific testing battery, ensuring athletes are judged across all the relevant fitness capacities that best define the physical demands of competition. To aid readers in using the TSA, this article also details how it is computed in EXCEL.

INTRODUCTION

In strength and conditioning coaches, we routinely put our athletes through a variety of fitness assessments to determine their

physical capability, so that we can tailor the design of their training program and adapt accordingly. Similarly, the physiotherapist, physiotherapist, and technical coaches also assess the athlete, with the results equally used to inform future interventions and team selection. But, with so much data collected and thus available for discussion, athlete review meetings, for example, where all staff attend, can often see each practitioner providing more discrete detail than is necessary. For example, although jump height may be informative to the strength and conditioning coach, this score, in this context, may not prove overly helpful to discussions led in to by the coaches and other members of the sport science disciplines. These situations, therefore, lend themselves to the strength and conditioning coach providing a single score for the athlete's physical fitness, rather than separately discussing each individual test result. Such an approach can streamline collaborative

communication, maximizing the time available for planning and practical delivery. Furthermore, coaches may not be as concerned in the raw score of each athlete, as much as where the score ranked among their teammates, especially when there is competition for places. For example, a coach may have no concept as to what is deemed a good jump height or back squat, with this information only becoming apparent through some analysis that reveals the score is among the highest or lowest in the squad. Also, it can be rare to have the athlete who scored highest on the bench press, also score the highest on a change of direction speed test or Yo-Yo score, for example, suggesting that there is some compromise among the different components of fitness that collectively

KEY WORDS: statistics; metrics; Excel; data analysis; testing; feedback

Copyright © National Strength and Conditioning Association. Strength and Conditioning Journal | www.nsc.org 01
 Copyright © National Strength and Conditioning Association. Unauthorized reproduction of this article is prohibited.

So, are they fit or not!?

But is that score any good!?

Are they getting better!?

COACH

Is that score any good and which test did they do best on?



- But maybe the team is fit and they all scored well on the shuttle test...
- Level 15 may have been one of the lowest
- Conversely, there may only be a few strong athletes, so 140kg is really good!

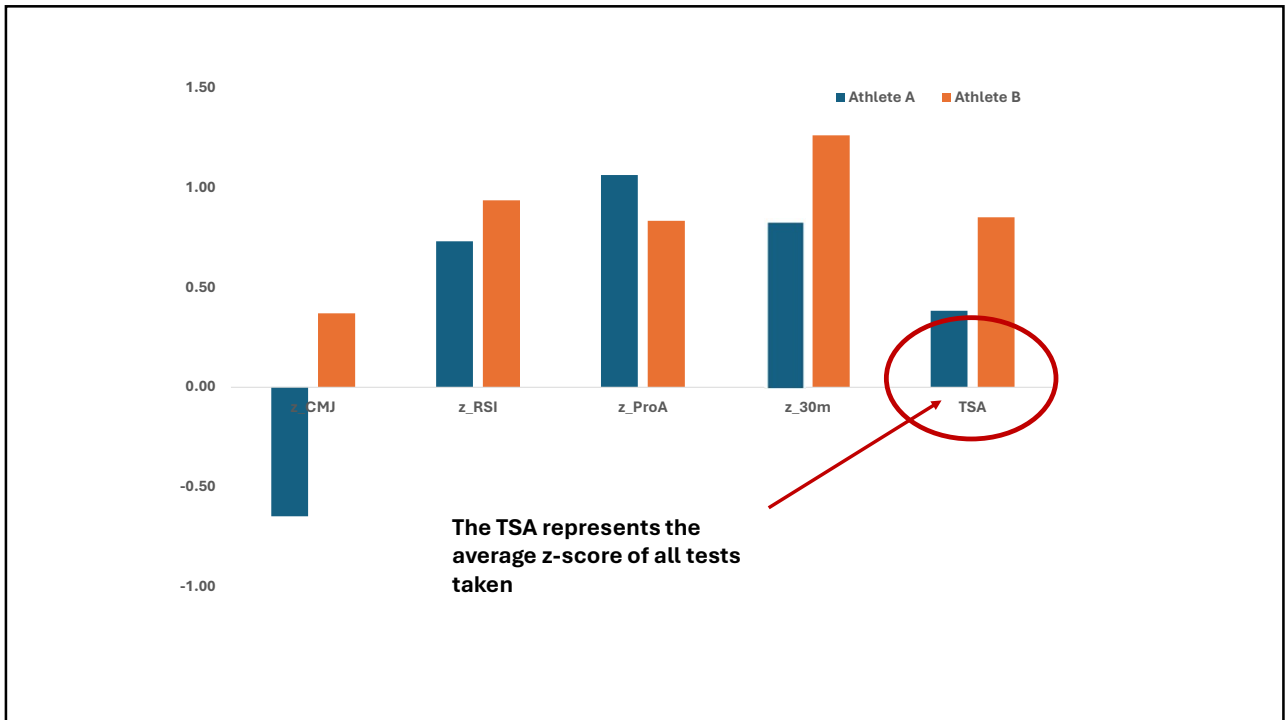
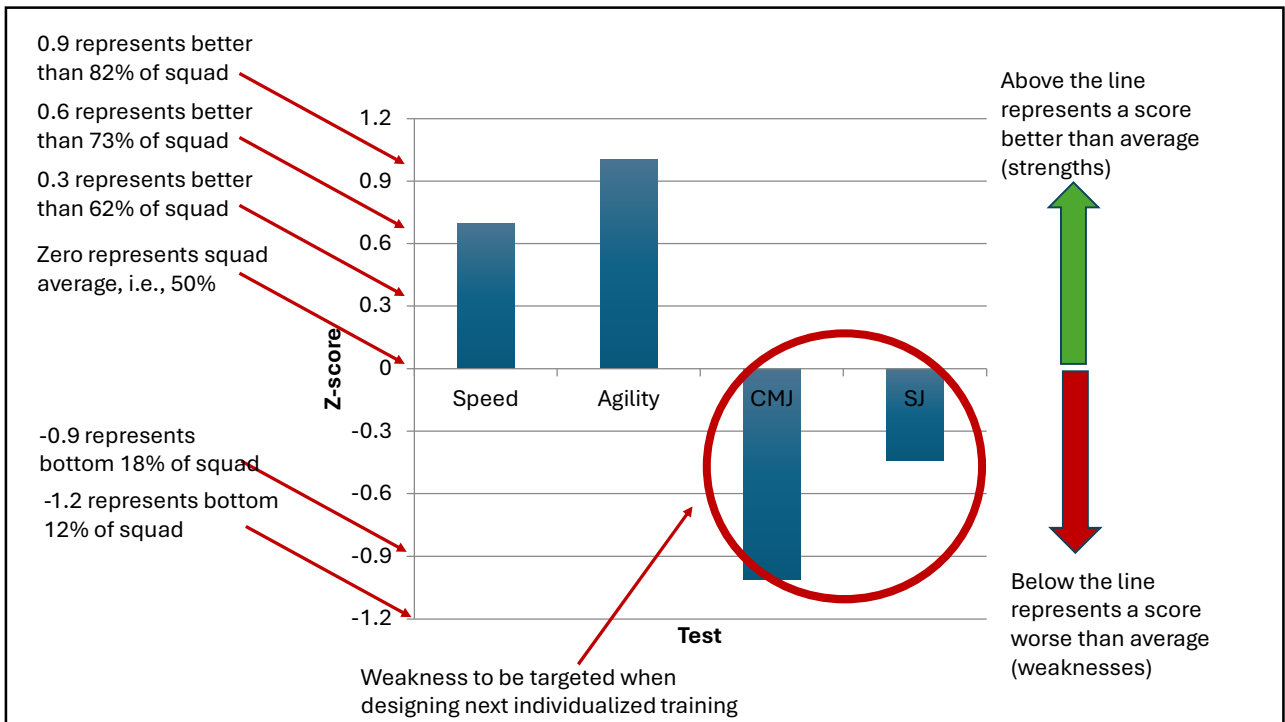
Turn test scores into a z-scores

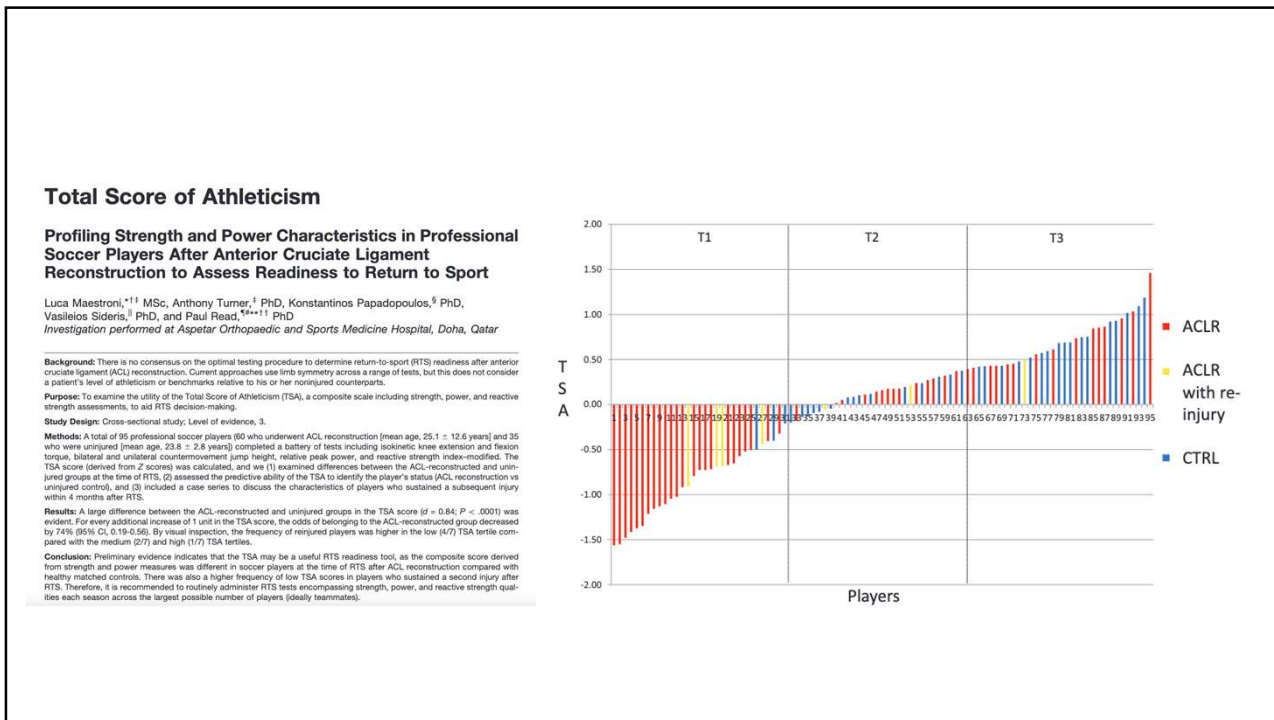
- Z-scores tell you how many SD's a score is from the mean
- If a z-score = 0, it is identical to the mean score
- If a z-score = 1, it is 1 SD above the mean
- If a z-score = -1, it is 1 SD below the mean

Then average them across all relevant tests

VLOOKUP fx =(S2-\$S\$26)/\$S\$27

	S	T	U	V	AE	AF	AG	AH	AI	AJ
1	Best_CMJ	Best_RSI	Best_ProA	Best_30m	z_CMJ	z_RSI	z_ProA	z_30m	TSA	Rank
2	47.7	1.6	5	4.5	SS27	-0.23	0.15	-1.49	-0.27	18
3	36	1.63	4.9	4.46	-0.77	-0.16	0.38	-1.25	-0.45	19
4	42.8	1.72	6.4	4.19	0.06	0.05	-3.06	0.41	-0.64	20
5	52.2	2.5	4.4	4.34	1.21	1.83	1.52	-0.51	1.01	2
6	37	2.02	4.6	4.25	-0.65	0.73	1.06	0.04	0.30	10
7	45.1	1.81	4.8	4.16	0.34	0.25	0.60	0.59	0.45	7
26	42.29	1.70	5.1	4.26						
27	8.16	0.44	0.44	0.16						





[Resources](#)
[Workshops](#)
[Research Papers](#)

Workshop Resources

★★★★★
Rated 5 stars by readers

Unlock Your Fitness Potential Today

Discover the science behind effective exercise programming and enhance your physical performance.

Determining Meaningful Change in Performance and Readiness When N = 1

Download the worksheet we will use to calculate the Coefficient of Variation as well as meaningful differences in performance (at 95% confidence) and readiness (athlete stability) testing.

[Download here](#)

www.TheFitnessFormula.Training/Workshops