Week 5 — Understanding measures

How well does the field establish, report, and interpret the validity and reliability of the measures its conclusions depend on?

Objective evidence review · companion to Week 5 · prepared July 2026

Scope & method. Rapid narrative review, not a systematic review. Structured search of the sport-science measurement and wearable-technology literature (PubMed, Google Scholar, European Journal of Applied Physiology, BMC). Findings reported as the cited studies state them; one adjacent-field source (measurement practices in psychology) is flagged.

The principle under review

WEEK 5 · VALIDITY, RELIABILITY, AND THE NOISE FLOOR

Week 5 argues that every measure must be assessed on two independent questions — does it measure the intended construct (validity) and does it give the same value on repeat (reliability) — that a change smaller than a measure's typical error is invisible, and that a device's reliability does not transfer automatically to a study's protocol. The empirical questions are whether the field establishes and reports these properties and interprets reliability indices correctly.

1 · How well is it currently applied?

THE EVIDENCE

43%	>20%	QMP
of studies reported the validity and/or reliability of the wearable devices they used — Rebelo et al. 2023	measurement error can coexist with an "excellent" ICC ≥ 0.90 — Warneke et al. 2025	measures often created "on the fly" without reported validity evidence — Flake & Fried 2020

Reliability indices are frequently misused and misinterpreted

A recent critical review with data simulations (Warneke, Gronwald, Wallot, Magno, Hillebrecht & Wirth, 2025) shows that the field's reliability on correlation-based reliability indices is often misleading: an "excellent" intraclass correlation (ICC ≥ 0.90) can coexist with measurement error exceeding 20%, because the ICC does not separate systematic error (learning, fatigue) from random error. The authors also document a common conflation of *device* reliability with *protocol* reliability — studies cite an instrument's reliability from elsewhere without establishing it for their own procedures, populations and settings — and the presentation of Bland–Altman limits of agreement without any pre-specified acceptable threshold. Each practice can make an unreliable measurement appear dependable.

Validity and reliability are often not reported at all

Even basic reporting is inconsistent. In a scoping review of wearable technologies used to inform injury-prevention strategies, Rebelo, Martinho, Valente-dos-Santos and Coelho-e-Silva (2023) found that only about 43% of studies reported the validity and/or reliability of the devices they used. This matters because commercial devices frequently rely on proprietary, unpublished algorithms, and their accuracy typically degrades under the dynamic, high-intensity conditions of actual sport relative to controlled or resting conditions.

Constructs are measured without established validity

The problem extends from hardware to constructs. Flake and Fried (2020) document widespread questionable measurement practices in which researchers create or adapt measures "on the fly" without justifying the choice or reporting validity evidence, and then treat a score as if it were the construct itself. Because "neither rigorous research design, nor advanced statistics, nor large samples can correct such false inferences," measurement weakness places a ceiling on the validity of every downstream conclusion.

The standards and tools exist and are established

The corrective methods are well established in the field. Hopkins (2000) set out how to quantify a measure's typical (random) error and its smallest worthwhile change; Atkinson and Nevill (1998) detailed the distinction between systematic bias and random error and appropriate reliability statistics; and Impellizzeri and Marcora (2009) adapted clinimetric validation standards to sport physiology, showing how surrogate and field measures should be validated before use. The gap is between these available standards and routine practice, not a lack of methods.

Source	Finding	Implication
Warneke et al. (2025) *Eur. J. Appl. Physiol.*	ICC ≥ 0.90 can hide >20% error; systematic vs random error conflated; device vs protocol reliability confused	Reliability is routinely over-stated and misinterpreted
Rebelo, Martinho, Valente-dos-Santos & Coelho-e-Silva (2023) *BMC Sports Sci. Med. Rehabil.*	Only ~43% of studies reported device validity/reliability	Measurement properties frequently unreported
Flake & Fried (2020) *AMPPS (adjacent)*	Measures created without justification or validity evidence; scores treated as constructs	Construct validity is often unestablished
Hopkins (2000) *Sports Medicine*	Framework for typical error and smallest worthwhile change	Method exists to define the noise floor
Atkinson & Nevill (1998) *Sports Medicine*	Systematic bias vs random error; appropriate reliability statistics	Method exists to report error correctly
Impellizzeri & Marcora (2009) *IJSPP*	Clinimetric validation standards for sport measures	Method exists to validate measures

2 · Is there cause for concern?

OBJECTIVE ASSESSMENT

ASSESSMENT **Moderate-to-substantial concern**

Two distinct problems combine. First, measurement properties are frequently not reported — fewer than half of wearable studies in one review reported device validity or reliability. Second, when reliability is reported, it is often expressed in indices (such as the ICC) that can conceal substantial error and that conflate systematic with random error, or borrowed from other studies rather than established for the protocol in hand. Because measurement is upstream of design, analysis and interpretation, unreliable or unvalidated measures place a ceiling on the trustworthiness of the conclusions — a measure whose error exceeds the effect of interest cannot detect that effect however well the rest of the study is conducted.

The concern is tempered by the fact that the corrective standards are long-established and sport-specific (Hopkins; Atkinson & Nevill; Impellizzeri & Marcora), and that recent critical work (Warneke et al., 2025) is actively raising the standard of reliability reporting. The problem is one of routine practice rather than missing methodology, and is correctable at the level of study design and editorial expectation.

3 · Immediate steps to improve

ACTIONABLE

1. **Report both validity and reliability of every key measure,** and cite the validation evidence for treating the measure as an index of the intended construct.
2. **Establish reliability for your own protocol** (population, equipment, procedures), rather than citing a device's reliability from another study.
3. **Report the typical (random) error and any systematic bias separately,** not a single correlation-based index; do not rely on the ICC alone.
4. **Pre-specify an acceptable error threshold and the smallest worthwhile change,** and check that the measure's error is below the effect you intend to detect (the noise floor).
5. **Be cautious with commercial/wearable metrics** whose algorithms are proprietary or unvalidated under match-intensity conditions.

References

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine, 26*(4), 217–238. https://doi.org/10.2165/00007256-199826040-00002

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science, 3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine, 30*(1), 1–15. https://doi.org/10.2165/00007256-200030010-00001

Impellizzeri, F. M., & Marcora, S. M. (2009). Test validation in sport physiology: lessons learned from clinimetrics. *International Journal of Sports Physiology and Performance, 4*(2), 269–277. https://doi.org/10.1123/ijspp.4.2.269

Rebelo, A., Martinho, D. V., Valente-dos-Santos, J., & Coelho-e-Silva, M. J. (2023). From data to action: a scoping review of wearable technologies and biomechanical assessments informing injury prevention strategies in sport. *BMC Sports Science, Medicine and Rehabilitation, 15*, 169. https://doi.org/10.1186/s13102-023-00783-4

Warneke, K., Gronwald, T., Wallot, S., Magno, A., Hillebrecht, M., & Wirth, K. (2025). Discussion on the validity of commonly used reliability indices in sports medicine and exercise science: a critical review with data simulations. *European Journal of Applied Physiology, 125*(6), 1511–1526. https://doi.org/10.1007/s00421-025-05720-6