



## OPEN ACCESS

## EDITED BY

Kevin Till,  
Leeds Beckett University, United Kingdom

## REVIEWED BY

Chris J. Bishop,  
Middlesex University, United Kingdom  
Franck Brocherie,  
Institut national du sport, de l'expertise et de la  
performance (INSEP), France

## \*CORRESPONDENCE

Lachlan P. James  
✉ l.james@latrobe.edu.au

RECEIVED 26 March 2024

ACCEPTED 03 June 2024

PUBLISHED 01 July 2024

## CITATION

James LP, Haycraft JAZ, Carey DL and  
Robertson SJ (2024) A framework for test  
measurement selection in athlete physical  
preparation.

Front. Sports Act. Living 6:1406997.

doi: 10.3389/fspor.2024.1406997

## COPYRIGHT

© 2024 James, Haycraft, Carey and  
Robertson. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# A framework for test measurement selection in athlete physical preparation

Lachlan P. James<sup>1\*</sup>, Jade A. Z. Haycraft<sup>1,2</sup>, David L. Carey<sup>1</sup> and Samuel J. Robertson<sup>2</sup>

<sup>1</sup>Sport, Performance, and Nutrition Research Group, School of Allied Health, Human Services, & Sport, La Trobe University, Melbourne, VIC, Australia, <sup>2</sup>Institute for Health and Sport, Victoria University, Footscray, VIC, Australia

Preparing athletes for competition requires the diagnosis and monitoring of relevant physical qualities (e.g., strength, power, speed, endurance characteristics). Decisions regarding test selection that attempt to measure these physical attributes are fundamental to the training process yet are complicated by the myriad of tests and measurements available. This article presents an evidenced based process to inform test measurement selection for the physical preparation of athletes. We describe a method for incorporating multiple layers of validity to link test measurement to competition outcome. This is followed by a framework by which to evaluate the suitability of test measurements based on contemporary validity theory that considers technical, decision-making, and organisational factors. Example applications of the framework are described to demonstrate its utility in different settings. The systems presented here will assist in distilling the range of measurements available into those most likely to have the greatest impact on competition performance.

## KEYWORDS

sport, assessment, validity, determinants, strength and conditioning

## 1 Introduction

The physical preparation of athletes requires a determination and assessment of relevant physical qualities (e.g., strength, power, speed, endurance characteristics) to identify strengths or weakness, and inform training interventions (1). Decisions regarding test selection that seek to measure these physical qualities are therefore a critical part of the training process (2). This process is increasingly more complicated given the large volume of tests and outputted measures available to performance staff. As such, a systematic decision-making framework to select context-specific physical testing measures that guide athlete preparation for sport is needed for practitioners (e.g., strength and conditioning coaches, sport scientists, performance analysts, and applied researchers). Conceptualization of validity as a network of inferences (3) and through measurement, organizational and decision-making perspectives can provide guidelines to practitioners when navigating the measurement selection process (4). However, a solution has yet to be presented in the literature in the context of athlete assessment models. As such, this article will use a contemporary understanding of validity to build a framework to guide test measurement selection for the physical preparation of athletes, that considers technical, decision-making, and organizational factors.

## 2 Validity

Validity is a primary consideration in test measurement selection as it considers the degree to which a test or measure is capable of achieving specific aims. Traditionally, three forms of validity were recognized: content, criterion, and construct (5). However, in the past 50 years, numerous forms of validity have been described in the literature, with definitions changing over this time (6). This has evolved further into a unified theory which places less emphasis on delineating dozens of types of validity, in favor of the few original forms and the accumulation of evidence for and against a validity proposition and situational context (4). In the context of sports performance, the forms of validity used to determine the relevance of a particular test measurement to sport performance is dependent on the sport and performance context. In measured sports (i.e., timed, weight, distance), for which deterministic models can be generated (7), criterion validity may be readily applied and simply assessed to establish this link. For example, a measure of barbell velocity during a power clean at a fixed load may be able to provide an accurate indication of weightlifting performance during competition. However, most sports can be classified as non-deterministic or complex, where competition outcomes are a multifactorial construct of events (e.g., team, field, court, and combat sports) (8), as such, criterion validity of a physical performance metric may not be easily assessed. In non-measured sports, content validity can refer to the degree to which the measurements included in a performance testing battery quantify the intended physiological characteristics that are important for sport performance (9). However, a crucial step in the validation process is to determine the extent to which physical performance qualities contribute to sport performance outcomes. Further, an analysis is required to identify which of these qualities are most relevant in a given context. In these cases, construct validity is a key lens through which to view validity, where the goal is to establish the relevance of a physical quality to sport performance. To represent this complex process, construct validity of physical performance metrics may be represented as a series of inferences that link from a test score to the performance outcome of interest.

## 3 Layers of validity

A network of inferences (3) or “layers of validity” can address the complexity of establishing measurement validity to the physical preparation process (Figure 1). This idea acknowledges that it may not be possible to establish a direct or causal relationship between a physical fitness measurement and successful competition performance outcomes (constructs) due to the complexity or abstract nature of sport competition. Instead, the relationship between physical performance measurements and competition outcomes may be assessed via intermediate steps or proxies that ultimately connect it to a competition performance outcome. For example, it may not be possible to determine whether a higher 1-RM back squat score leads to more wins in professional rugby

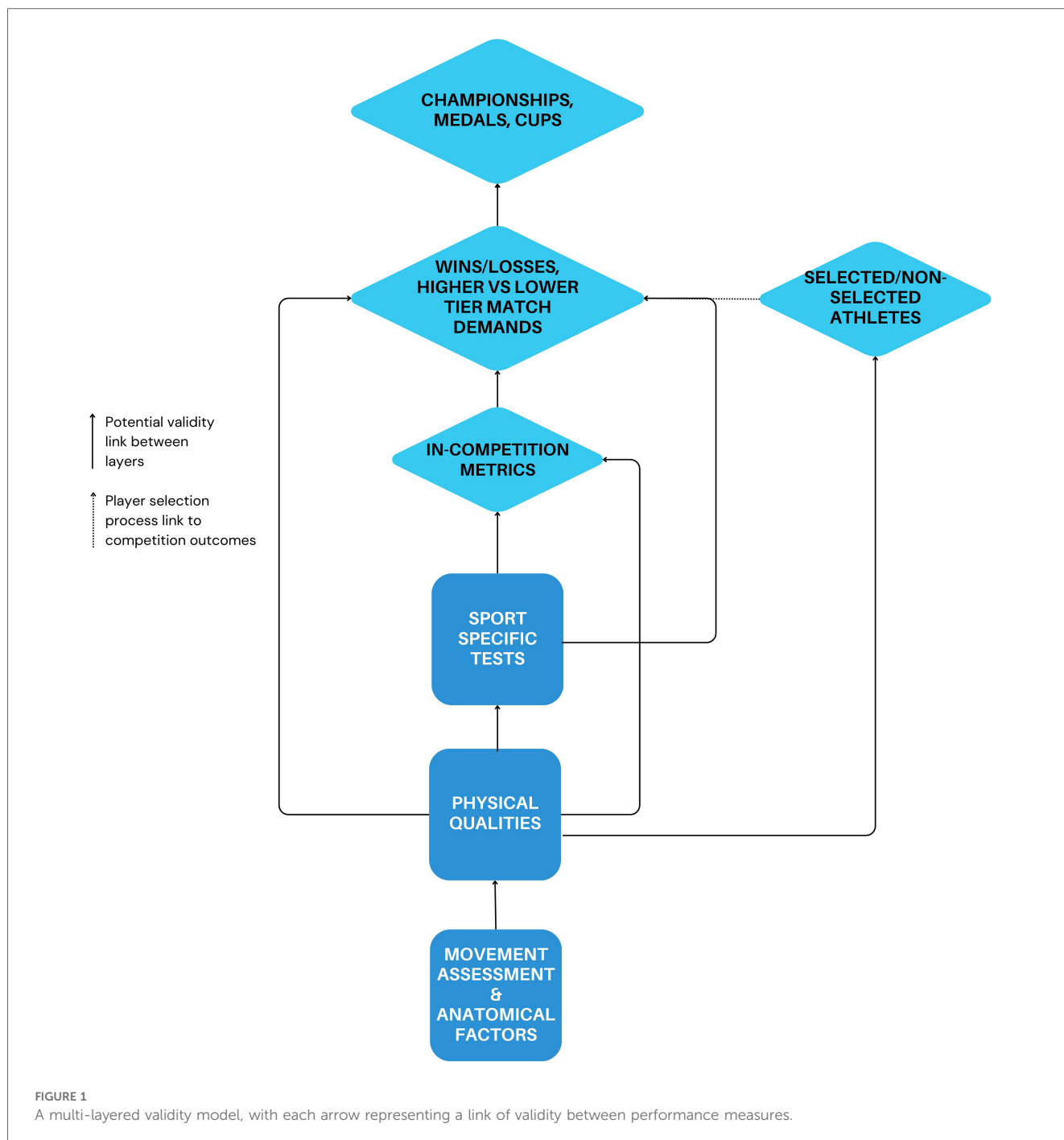
league, but instead may be associated with within-game measures of tackling ability (10), which are in turn characteristic of higher- vs. lower-tier match-play (11). Furthermore, absolute 1-RM back performance distinguishes professional from semi-professional rugby league players to a greater extent than alternate metrics (12) (Figure 2). In this case, multiple layers of evidence are used to bridge the gap between the physical performance measurements and the construct of successful competition performance outcomes.

The process of establishing layers of validity involves breaking down the relationship between the physical performance measurements and the competition performance outcomes into a series of smaller, more manageable relationships or steps, each with its own validity evidence. By demonstrating the validity of these intermediate steps, the overall validity of the measure with respect to the intended construct can be supported indirectly. As the number of intermediate steps between the metric and the competition outcome increase so too does the complexity of the argument (13). However, this process allows for the accumulation of evidence towards the relationships between a measure and the sport performance outcome (14). Although a multi-layered validity model may differ somewhat based on the context, the objective is to provide a schematic of measurement-performance interaction that guides the analytical approach for constructing a validity argument. Therefore, it is also necessary to have a complementary framework to determine the *extent to which* one layer is connected to another.

## 4 Physical assessment framework

The objective of the framework is to evaluate the ability of a given test measurement to link the steps within the multi-layered validity model. The items listed within the framework are informed by contemporary validity theory (4) and therefore considers three primary criteria: (i) measurement (Table 1), (ii) decision-making (Table 2), and (iii) organizational (Table 3). The measurement criteria assess how a metric is collected, the quality and type of evidence supporting connections to performance outcomes, and the amount of unique information it captures (5). Decision-making is captured in the second criteria and considers the extent decisions can be made based upon the measurement (15). The organizational criteria accounts for the feasibility of implementing the test and measurement within the sport organization (16). The primary objectives of this framework are supplemented with secondary criteria that identify future side-effects or consequences (positive or negative) of utilizing the measure of interest that extend beyond the identified primary objectives (4). Also informing the framework are previous models developed to guide the implementation of sports technology (17, 18), and jump assessments (19), across contexts. Taken together, the proposed framework provides a systematic decision support process for sports practitioners that incorporates the technical and organizational needs of measurement selection for the physical preparation of athletes.





## 4.1 Measurement criteria

### 4.1.1 Measurement collection

A metric must be viewed in the context of how it is captured, so the validity and reliability of the collection method is an essential consideration (20, 21). For example, linear position transducers have been shown to be a more valid and reliable method to measure barbell velocity during resistance training than accelerometer based devices (22). Metrics without evidence of reliability have limited utility for decision making within a validity model since variation may come from sources external to

the athlete, such as measurement error, the identity of the person running the test, or the testing environment (23). Knowledge of the measurement error for a test is important to contextualize the size of any observed changes, and to avoid over-interpretation of individual changes that are small relative to the measurement error. Without validity it is difficult to understand the mechanisms of action between layers in the validity model and thus make targeted interventions. For example, an invalid test of change of direction ability may still show associations with performance outcomes if the majority of the variance is explained by a different physical property (such as speed) (24),

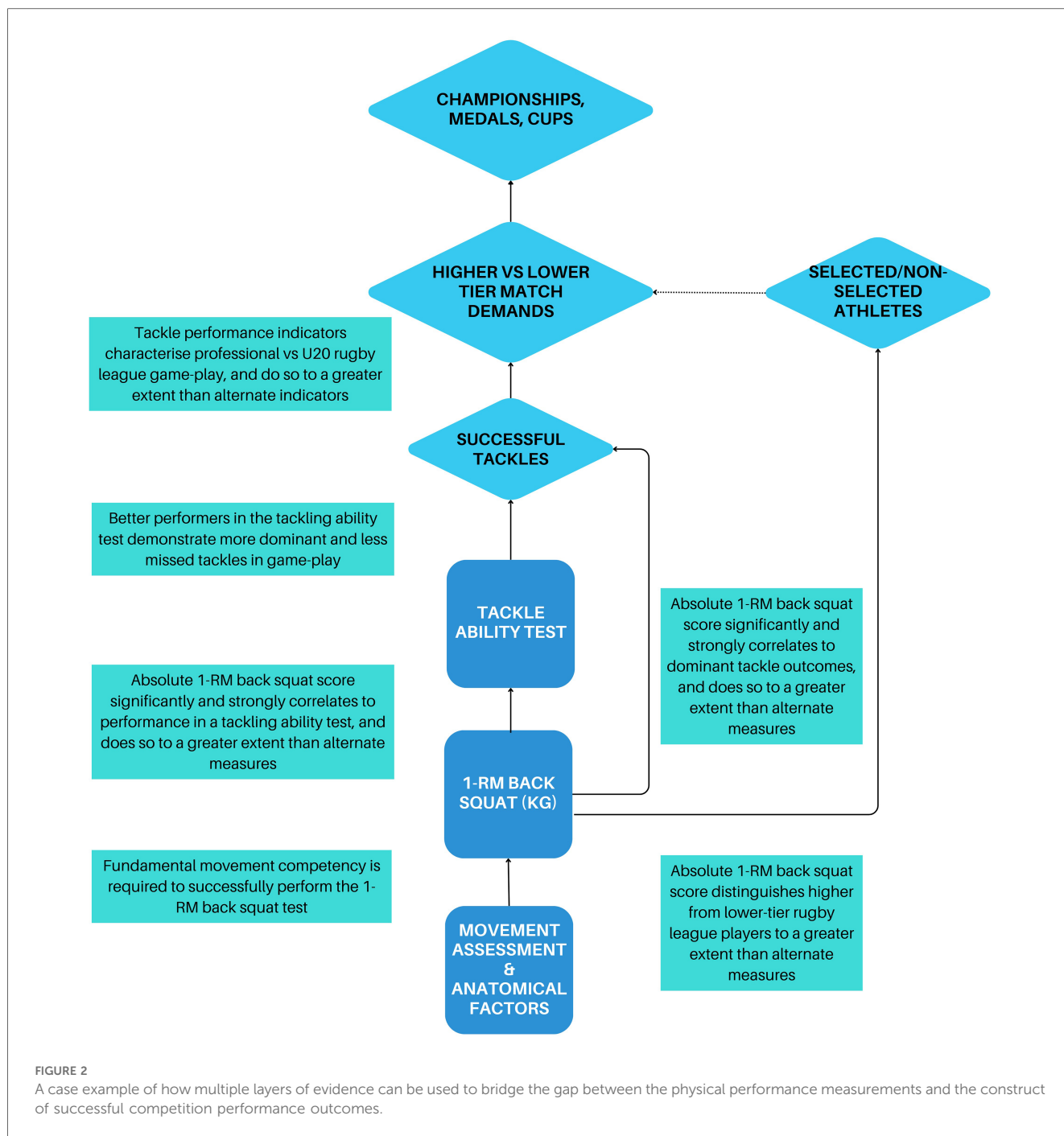


TABLE 1 Measurement criteria.

Measurement criteria	Definition	Scale
Metric collection	How valid and reliable is the proposed method of metric collection?	Low/Moderate/Strong
Evidence of association	How much, and what kind of evidence is there for a proposed link between a test metric and a performance outcome? Is there a plausible mechanistic explanation grounded in domain knowledge?	Low/Moderate/Strong
Nature of association	Has the type of association been explored?	Low/Moderate/Strong
Independence of information	Is the proposed metric measuring something already captured by another metric in the model?	No/Somewhat/Yes

TABLE 2 Decision making criteria.

Decision making criteria	Definition	Scale
Interpretability	Interpretability of physical performance metrics is the degree of meaningful insights that can be extracted from the metric, or a change in the metrics value can be used by coaches, sport science staff, and broader organization stakeholders to inform decision-making.	Low/Moderate/Strong
Responsiveness	The ability for a measurement to detect real changes to strength and conditioning interventions over time. Question that may be asked by the practitioner are: Can the measure inform training interventions? Is the measured attribute modifiable?	Low/Moderate/Strong
Diminishing returns	As a person receives more exposure to a given stimulus, their ability to adapt generally diminishes. Training may therefore be directed to an alternate, but still relevant physical quality. How well does the measure consider this criterion?	Low/Moderate/Strong

TABLE 3 Organizational criteria.

Organisational Criteria	Definition	Scale
Financial cost	The financial cost is as any monetary cost associated with physical performance measurements. This may include, but is not limited to, purchase and maintenance of equipment, venue cost and overheads (i.e., utility costs, internet), and staff wages for data collection, processing, analysis, and reporting. The financial costs of physical performance measures should remain within the budgetary constraints of the sports organization or governing bodies (i.e., soft capping of team finances, government sport funding)	Low cost, moderate cost, high cost
Opportunity cost	The opportunity cost of implementing a physical performance metric test are resources that may need to be sacrificed from other areas of sport program to allow a specific physical performance metric to be implemented. Opportunity costs may include the re-allocation of financial costs from other departments (i.e., nutrition, sport psychology, analytics), or sacrificing of time from training sessions (i.e., match-simulation, skills sessions, sport-specific training) to accommodate the physical performance testing session	Low cost, moderate cost, high cost
Time cost of test familiarization	The time cost of test familiarization considers the test complexity when implementing and administering a test. Increased task complexity within performance tests may require higher levels of expertise which incur a time cost in upskilling staff (i.e., equipment, software, pass/fail criteria) and athletes (i.e., pacing strategies, coordination of multi-joint complex movements) on test methods	Low cost, moderate cost, high cost
Time-cost of implementing in the training environment	Time-cost of test implementation considers timing and duration of physical performance testing within a training session, week, annual or bi-annual plan. Within the time-cost of test implementation, recovery time cost is also considered athletes to mitigate risk of injury or overtraining from residual fatigue between tests, testing days, and training days	Low cost, moderate cost, high cost

however subsequent decision-making regarding training interventions may be targeted at the wrong physical qualities. This consideration has become acutely important with advances in technology and over-saturation in the market of measurement devices in sport (17).

#### 4.1.2 Evidence of association

Assessment of the strength of evidence supporting a connection between a test metric and performance outcome is multi-faceted and often challenging (25). A connection should ideally be supported by both a plausible mechanistic explanation and supporting experimental or generalizable observational data (26). Explanations grounded in domain knowledge of physiology and sport performance are important for establishing causal links and for interpreting published data (26, 27), and considerations of causality are necessary for the design and implementation of training interventions. For example, a study of recreational and untrained cyclists reported a strong correlation ( $r=0.965$ ) between maximum power in an incremental test (cycling to exhaustion from 100 W, increasing at  $20 \text{ W}\cdot\text{min}^{-1}$ ) and functional threshold power (maximum average power sustained over a 1 h period) (28). Additionally, the study found that both the test value and functional threshold power significantly increased after a training intervention. In this example the practitioner can be confident that improvements in the

incremental test will translate to performance gains in sub-elite cyclists. However, the generalizability of this finding to elite cohorts is not guaranteed since the trainability of the underlying physical qualities may be different, highlighting the importance of appraising evidence of association relative to the cohort considered. In addition to the presence of a mechanistic explanation, the body of empirical evidence supporting a link should be appraised within the traditional sources of evidence hierarchy (29), with attention given the magnitude of effects interpreted in the context of the desired outcome (30). Study population characteristics are another important consideration, and preference may be given to selecting metrics with associations shown within the same sport specialization and level of proficiency.

#### 4.1.3 Nature of association

After establishing an association, further understanding of how a metric is related to a performance outcome can inform the construction of the multi-layered validity model. Aspects to look for in the evidence include methods that allow for non-linear relationships, interacting (31), mediating (32, 33), or moderating (34) effects. Analytical techniques such as machine learning algorithms can permit for this non-linearity and complexity in metrics to be uncovered, however the use of such analysis methods does not supplant the need for plausible mechanistic

and causal explanations (35). Examples include the identification of tipping points (e.g., the point between functional and non-functional overreaching), feedback loops (how an increase in maximal strength allows for enhanced adaptations to power training), or emergence (performance a live tester attacker vs. defender test can discriminate between higher and lower-level players, while a preplanned change of direction test cannot) (36). Having an awareness of these phenomena [see (36) for further reading on the topic], along with appropriate analytical tools to account for them can assist the strength and conditioning coach when selecting valid physical performance metrics within the team sport environment.

#### 4.1.4 Independence of information

In some cases, a set of multiple metrics may be assessed to have equal measurement validity, reliability, and association with performance, in these situations dimensionality reduction methods can assist selection (37, 38). Metrics that exhibit lower co-linearity with others in the proposed multi-layered validity model can be prioritized as containing more independent (less redundant) information. This is a desirable quality since it allows for clearer resolution when assessing an athlete's physical abilities if each metric represents a distinct quality (1).

### 4.2 Decision making criteria

Decision making objectives consider the extent to which a user can knowingly act on the information contained in the measurement (4, 39) (Table 2). A key feature of this is interpretability, which reflects the capacity to analyze and integrate physical performance measures into training and competition protocols (i.e., the practical utility of the measurement) (21). While decisions can be made in the absence of interpretation, they are more likely to lead to errors and unintended consequences. Test measurements should align with specific objectives, such as performance enhancement, talent identification, or injury prevention, and should inform the development of physical training programs. Therefore, the influence of specific decision making factors may change depending which aspect of the physical preparation plan is of interest. To strengthen the interpretability of these metrics, normalization within a team or comparative benchmarking against competitors or comparable sports can offer a contextual reference, resulting in improved utility of the data (1, 40). A performance metrics' responsiveness reflects its accuracy in detecting meaningful changes in response to a defined stimulus (41, 42). In physical preparation for sport, this stimulus is often characterized by training interventions, but can also be influenced by variables such as fatigue, adaptations from other training activities, or from competition itself. The end-user must also consider the principle of diminishing returns in the context of training stimulus and adaptation. Diminishing returns refers to the observation that as an athlete receives greater exposure to a stimulus, the rate of adaptation decreases, and alternate

training stimuli are required to achieve continual physical performance gains (43–45). Therefore, as an athlete's training progresses, the relevance of a specific physical performance metric may decrease due to a plateau in potential adaptation in one physical domain, meanwhile another physical performance metric may gain prominence as it allows for a greater potential for physical performance improvement.

### 4.3 Organisational criteria

Organisational objectives (Table 3) reflect the feasibility of implementing a physical performance testing and analysis process within the sports organization. Feasibility assessments are commonly used in sport business, management, strategic planning, and marketing to determine the costs vs. benefits of business strategies, processes, projects, programs, and facilities (46). The costing considerations of selecting physical performance metrics pertains to the financial cost, time cost, and opportunity cost of undertaking the testing and analysis (21, 47). The inclusion of a feasibility evaluation of a physical performance metric provides practitioners and their organizations information on the cost vs. benefit of their testing battery. For example, a feasibility evaluation may allow practitioners to quantify the value of a performance test purpose (i.e., training intervention, talent identification, injury prevention), financial costs (i.e., staff wages, equipment, venue), time cost (i.e., set-up time, time athletes take to complete test, data analysis and interpretation, staff upskilling), physical risk vs. benefit of testing (i.e., injury, adverse training/performance affects), and identification of similar performance metrics assessed across multiple tests (21, 48).

### 4.4 Secondary criteria

Secondary criteria represent impacts that follow the implementation of a measure that are outside of the primary objective (4). This might include factors such as athlete motivation, feedback, teamwork, competition. Secondary objectives also consider whether the inclusion of a physical performance metric has a wider impact on the organization or sport beyond the performance department such as its ability to aid developmental pathways, or foster grassroots sport. Negative consequences can also be considered as a secondary objective. For example, a secondary objective assessment may identify potential scenarios where the physical performance metric to be used incorrectly, such as its use in the wrong sub populations. Errors leading to wrong decision-making can be costly for a sports organization as they result in wastage of resources, thus affecting the team's financial stability and may lead to budgetary constraints in other areas the sports organization. Secondary impacts may therefore range from internal financial implications within an organization, to those impacts that have a wider social or sports governance impact that is challenging to quantify.

**A**

		MEASURE 1	MEASURE 2	MEASURE 3
Measurement Objectives	Association between layers	✓	×	✓
	Metric collection	Strong		Moderate
	Evidence of association	Moderate		Low
	Nature of association	Strong		Moderate
	Independence of information	✓		×
Decision Making Objectives	Interpretability	Low		Strong
	Responsiveness	Moderate		Moderate
	Diminishing returns	Strong		Moderate
Organizational Objectives	Financial cost	High Cost		Low Cost
	Opportunity cost	High Cost		Moderate
	Time cost of tester familiarization	Moderate		Low Cost
	Time cost of implementation	High Cost		Low Cost

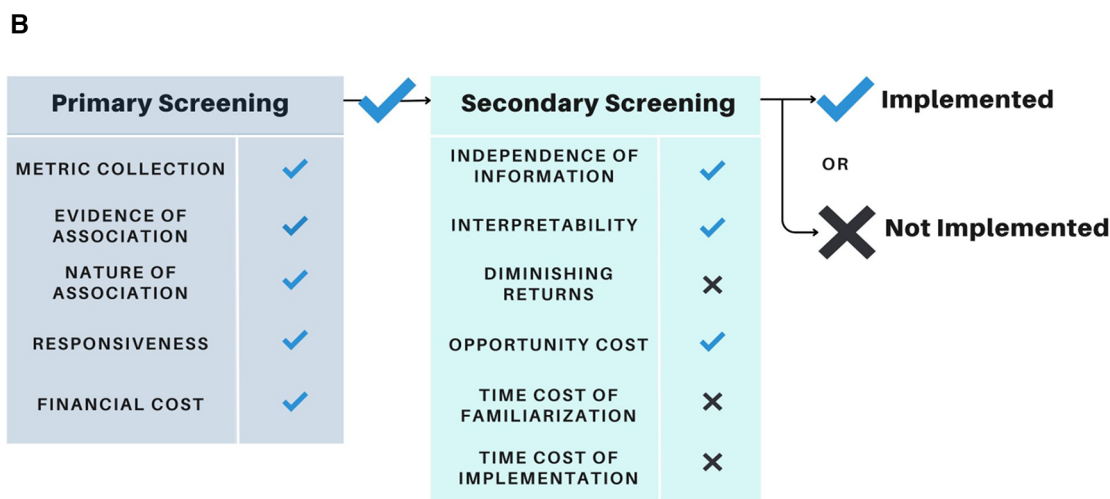


FIGURE 3

(A) A case example of how the framework can be applied via a checklist format. In this example, the tool commences with a single gatekeeper question, followed by each of the remaining criteria. The gatekeeper question states: "Is there evidence of association between the measure and another "layer" in the validity model?". Without a positive response to this question, the remaining items and objectives are of little value. Adapted from Robertson et al. (18). (B) A case example of how the framework can be applied with items weighted differently based on their perceived importance. In this example, a potential measurement must past the primary round of screening before further consideration is given. The secondary round of screening may not require all selected items to be met prior to the measurement's implementation into a testing battery. Adapted from Robertson et al. (18).



## 5 Framework application

The proposed physical assessment framework has been developed to allow flexibility in its application across various sport contexts. At the broadest level, it can represent a checklist by which to compare multiple measurements that are being considered for implementation (18). In this example (Figure 3A), the tool commences with a single gatekeeper question, followed by each of the remaining criteria. The gatekeeper question states: “Is there evidence of association between the measure and another “layer” in the validity model?”. Without a positive answer to this question, the remaining items and objectives are of little value. This provides a broad, non-prescriptive model that can be used across a range of situations.

The framework may also be tailored to meet the specific needs of a physical preparation program, training environment, sports organization, or governing body. This can be done by weighting the items within each objective category more or less heavily than other items via multiple rounds of screening (18). In the example provided (Figure 3B) several items have been selected by an end-user as they may be deemed more important than others in their specific use case. In this example, a potential measurement must pass the primary round of item and objective screening before further consideration is given. The secondary round of screening may not require all selected items to be met prior to the metric’s implementation. As such, the secondary round of testing may serve to identify any limitations or constraints of a physical performance metric if the end-user decides to implement it within their training program or organization.

## 6 Conclusion

This article presented evidenced based processes to inform test measurement selection for the physical preparation of athletes. We described a method for building a multi-layered validity model, followed by a framework by which to evaluate the suitability of metrics that considers technical, decision-making, and

organizational factors. The framework can be applied in several ways depending on the needs of the user, including a checklist design or a gatekeeper model. The systems presented here will assist in distilling the myriad of measurements available into those likely to have the greatest impact on performance.

## Author contributions

LJ: Conceptualization, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. JH: Methodology, Visualization, Writing – original draft, Writing – review & editing. DC: Methodology, Writing – original draft, Writing – review & editing. SR: Methodology, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- James LP, Talpey SW, Young WB, Geneau MC, Newton RU, Gastin PB. Strength classification and diagnosis: not all strength is created equal. *Strength Cond J.* (2023) 45(3):333–41. doi: 10.1519/SSC.0000000000000744
- Jeffries AC, Marcora SM, Coutts AJ, Wallace L, McCall A, Impellizzeri FM. Development of a revised conceptual framework of physical training for use in research and practice. *Sports Med.* (2022) 52(4):709–24. doi: 10.1007/s40279-021-01551-5
- Kane MT. Current concerns in validity theory. *J Educ Meas.* (2001) 38(4):319–42. doi: 10.1111/j.1745-3984.2001.tb01130.x
- Shaw SD, Newton PE. Twenty-First-Century evaluation. In: Newton PE, Shaw SD, editors. *Validity in Educational and Psychological Assessment.* London, UK: Sage (2014). p. 183–226.
- Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull.* (1955) 52(4):281. doi: 10.1037/h0040957
- Shaw SD, Newton PE. Validity and validation. In: Newton PE, Shaw SD, editors. *Validity in Educational and Psychological Assessment.* London, UK: Sage (2014). p. 1–25.
- Chow JW, Knudson DV. Use of deterministic models in sports and exercise biomechanics research. *Sports Biomechanics.* (2011) 10(3):219–33. doi: 10.1080/14763141.2011.592212
- Atkinson G. Sport performance: variable or construct? *J Sports Sci.* (2002) 20(4):291–2. doi: 10.1080/026404102753576053
- Gogos BJ, Larkin P, Haycraft JA, Collier NF, Robertson S. Combine performance, draft position and playing position are poor predictors of player career outcomes in the Australian football league. *PLoS One.* (2020) 15(6):e0234400. doi: 10.1371/journal.pone.0234400
- Speranza MJ, Gabbett TJ, Greene DA, Johnston RD, Townshend AD. Tackle characteristics and outcomes in match-play rugby league: the relationship with tackle ability and physical qualities. *Sci Med Footb.* (2017) 1(3):265–71. doi: 10.1080/24733938.2017.1361041
- Woods CT, Sinclair W, Robertson S. Explaining match outcome and ladder position in the national rugby league using team performance indicators. *J Sci Med Sport.* (2017) 20(12):1107–11. doi: 10.1016/j.jsams.2017.04.005

12. Baker DG, Newton RU. Comparison of lower body strength, power, acceleration, speed, agility, and sprint momentum to describe and compare playing rank among professional rugby league players. *J Strength Cond Res.* (2008b) 22(1):153–8. doi: 10.1519/JSC.0b013e31815f9519
13. Sirci SG. Agreeing on validity arguments. *J Educ Meas.* (2013) 50(1):99–104. doi: 10.1111/jedm.12005
14. Edwards MC, Slagle A, Rubright JD, Wirth R. Fit for purpose and modern validity theory in clinical outcomes assessment. *Qual Life Res.* (2018) 27:1711–20. doi: 10.1007/s11136-017-1644-z
15. Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas.* (2013) 50(1):1–73. doi: 10.1111/jedm.12000
16. Cronbach LJ, Ambron SR, Dornbusch SM, Hess RD, Hornik RC, Phillips DC, et al. *Toward Reform of Program Evaluation.* New York: Citeseer (1980).
17. Windt J, MacDonald K, Taylor D, Zumbo BD, Sporer BC, Martin DT. To tech or not to tech? A critical decision-making framework for implementing technology in sport. *J Athl Train.* (2020) 55(9):902–10. doi: 10.4085/1062-6050-0540.19
18. Robertson S, Zender J, De Mey K, Haycraft J, Ash G, Brockett C, et al. Development of a sports technology quality framework. *J Sports Sci.* (2024) 41:1–11. doi: 10.1080/02640414.2024.2308435
19. Bishop C, Turner A, Jordan M, Harry J, Loturco I, Lake J, et al. A framework to guide practitioners for selecting metrics during the countermovement and drop jump tests. *Strength Cond J.* (2022) 44(4):95–103. doi: 10.1519/ssc.0000000000000677
20. Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. *Sports Med.* (2008) 38(4):297–316. doi: 10.2165/00007256-200838040-00003
21. Robertson S, Kremer P, Aisbett B, Tran J, Cerin E. Consensus on measurement properties and feasibility of performance tests for the exercise and sport sciences: a delphi study. *Sports Med Open.* (2017) 3(1):2. doi: 10.1186/s40798-016-0071-y
22. Weakley J, Morrison M, García-Ramos A, Johnston R, James L, Cole MH. The validity and reliability of commercially available resistance training monitoring devices: a systematic review. *Sports Med.* (2021) 51:1–60. doi: 10.1007/s40279-020-01382-w
23. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* (1998) 26(4):217–38. doi: 10.2165/00007256-199826040-00002
24. Nimphius S, Callaghan SJ, Spiteri T, Lockie RG. Change of direction deficit: a more isolated measure of change of direction performance than total 505 time. *J Strength Cond Res.* (2016) 30(11):3024–32. doi: 10.1519/JSC.0000000000001421
25. Steele J, Fisher J, Crawford D. Does increasing an athletes' strength improve sports performance? A critical review with suggestions to help answer this, and other, causal questions in sport science. *J Trainol.* (2020) 9(1):20. doi: 10.17338/trainology.9.1\_20
26. Rohrer JM. Thinking clearly about correlations and causation: graphical causal models for observational data. *Adv Meth Pract Psychol Sci.* (2018) 1(1):27–42. doi: 10.1177/2515245917745629
27. Rommers N, Rössler R, Shrier I, Lenoir M, Witvrouw E, D'Hondt E, et al. Motor performance is not related to injury risk in growing elite-level male youth football players. A causal inference approach to injury risk assessment. *J Sci Med Sport.* (2021) 24(9):881–5. doi: 10.1016/j.jsams.2021.03.004
28. Denham J, Scott-Hamilton J, Hagstrom AD, Gray AJ. Cycling power outputs predict functional threshold power and maximum oxygen uptake. *J Strength Cond Res.* (2020) 34(12):3489–97. doi: 10.1519/JSC.0000000000002253
29. Ardern CL, Dupont G, Impellizzeri FM, O'Driscoll G, Reurink G, Lewin C, et al. Unravelling confusion in sports medicine and sports science practice: a systematic approach to using the best of research and practice-based evidence to make a quality decision. *Br J Sports Med.* (2019) 53(1):50–6. doi: 10.1136/bjsports-2016-097239
30. Caldwell A, Vigotsky AD. A case against default effect sizes in sport and exercise science. *PeerJ.* (2020) 8:e10314. doi: 10.7717/peerj.10314
31. Newans T, Bellinger P, Drovandi C, Buxton S, Minahan C. The utility of mixed models in sport science: a call for further adoption in longitudinal data sets. *Int J Sports Physiol Perform.* (2022) 1(aop):1–7. doi: 10.1123/ijssp.2021-0496
32. Mooney M, O'Brien B, Cormack S, Coutts A, Berry J, Young W. The relationship between physical capacity and match performance in elite Australian football: a mediation approach. *J Sci Med Sport.* (2011) 14(5):447–52. doi: 10.1016/j.jsams.2011.03.010
33. Domaradzki J, Popowczak M, Zwierko T. The mediating effect of change of direction speed in the relationship between the type of sport and reactive agility in elite female team-sport athletes. *J Sports Sci Med.* (2021) 20(4):699. doi: 10.52082/jssm.2021.699
34. Vaughan RS, Laborde S. Attention, working-memory control, working-memory capacity, and sport performance: the moderating role of athletic expertise. *Eur J Sport Sci.* (2021) 21(2):240–9. doi: 10.1080/17461391.2020.1739143
35. Brocherie F, Chassard T, Toussaint J-F, Sedeaud A. Comment on: "black box prediction methods in sports medicine deserve a red card for reckless practice: a change of tactics is needed to advance athlete care". *Sports Med.* (2022) 52(11):2797–8. doi: 10.1007/s40279-022-01736-6
36. Boehnert J. *The Visual Representation of Complexity: Sixteen Key Characteristics of Complex Systems* (2018).
37. James LP, Suppiah H, McGuigan M, Carey DL. Dimensionality reduction for countermovement jump metrics. *Int J Sports Physiol Perform.* (2021) 16(7):1052–5. doi: 10.1123/ijssp.2020-0606
38. Merrigan JJ, Stone JD, Ramadan J, Hagen JA, Thompson AG. Dimensionality reduction differentiates sensitive force-time characteristics from loaded and unloaded conditions throughout competitive military training. *Sustainability.* (2021) 13(11):6105. doi: 10.3390/su13116105
39. Guion RM. Recruiting, selection, and job placement. *Handbook of Industrial and Organizational Psychology.* New York: Rand McNally College Pub. Co. (1976). p. 777–828.
40. McMahon JJ, Lake JP, Comfort P. Identifying and reporting position-specific countermovement jump outcome and phase characteristics within rugby league. *PLoS One.* (2022) 17(3):e0265999. doi: 10.1371/journal.pone.0265999
41. Beaton D, Bombardier C, Katz J, Wright J. A taxonomy for responsiveness. *J Clin Epidemiol.* (2001) 54(12):1204–17. doi: 10.1016/S0895-4356(01)00407-3
42. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and use.* Oxford: Oxford university press (2024).
43. Newton RU, Kraemer WJ. Developing explosive muscular power: implications for a mixed methods training strategy. *Strength Cond J.* (1994) 16:20–31. doi: 10.1519/1073-6840(1994)016<0020:DEMPPI>2.3.CO;2
44. Baker D, Newton R. Observation of 4-year adaptations in lower body maximal strength and power output in professional rugby league players. *J Aust Strength Cond.* (2008a) 16(1):3–10.
45. Appleby B, Newton RU, Cormie P. Changes in strength over a 2-year period in professional rugby union players. *J Strength Cond Res.* (2012) 26(9):2538–46. doi: 10.1519/JSC.0b013e31823f8b86
46. Ali A. Measuring soccer skill performance: a review. *Scand J Med Sci Sports.* (2011) 21(2):170–83. doi: 10.1111/j.1600-0838.2010.01256.x
47. Gabbett HT, Windt J, Gabbett TJ. Cost-benefit analysis underlies training decisions in elite sport. *Br J Sport Med.* (2016) 50:1291–2. doi: 10.1136/bjsports-2016-096079
48. Hughes MD, Bartlett RM. The use of performance indicators in performance analysis. *J Sports Sci.* (2002) 20(10):739–54. doi: 10.1080/026404102320675602