Testing and Profiling Athletes: Recommendations for Test Selection, Implementation, and Maximizing Information

Abstract

Understanding the physical qualities of athletes can lead to improved training prescription, monitoring, and ranking. Consequently, testing and profiling athletes is an important aspect of strength and conditioning. However, results can often be difficult to interpret due to the wide range of available tests and outcome variables, the diverse forms of technology used, and the varying levels of standardization implemented. Furthermore, physical qualities can easily be misrepresented without careful consideration if fundamental scientific principles are not followed. This review discusses how to develop impactful testing batteries so that practitioners can maximize their understanding of athletic development while helping to monitor changes in performance to better individualize and support training. It also provides recommendations on the selection of tests and their outcome measures, considerations for the proper interpretation, set-up, and standardization of testing protocols, methods to maximize testing information, and techniques to enhance visualization and interpretation.

Keywords: Physical Qualities; Monitoring; S&C; Technology; Coaching; Strength and Conditioning

Introduction

The testing and profiling of athletes are essential for strength and conditioning coaches. Data from carefully constructed testing batteries can ensure a competitive edge over the opposition by providing information to better guide training prescription and monitor changes in performance (58). Furthermore, information gleaned from testing can be used to identify talent and help justify the selection of athletes (16, 36, 72, 81). However, testing can also be misused, resulting in physical qualities being misunderstood or misrepresented (34, 51). Therefore, if information is being gathered to help guide the decisions of coaches, it is important to ensure that the most accurate and impactful information is being collected and presented. This is particularly important for teams or sporting organizations investing significant time and resources into an athlete.

Considering the importance of testing for coaches and athletes, it is essential to consider *why* and *how* the testing is being implemented. While the growing acceptance of sport science and technology has helped to continue the development and innovation within strength and conditioning (85, 99), it has also led to extremely large amounts of data often being available (56). This can cause practitioners to be overwhelmed with information (i.e., "paralysis through analysis"), select inappropriate testing methods or outcomes (i.e., the lack of understanding of the test and its underpinning physiological/biomechanical constructs), or cause 'testing for testing's sake'. Thus, understanding the '*why*' can support decisions around what information is retained and help determine the purpose, which in turn can help guide the tests that are selected. Furthermore, once the tests have been decided upon, '*how*' testing occurs is essential to establish as this ensures the integrity of the retrieved information. *How* testing is conducted can make a substantial difference to the outcomes of nearly all tests and encompasses how tests are standardized and implemented, the equipment and variables used, and how the data is handled.

With physical testing being an integral part of strength and conditioning, it is important to acknowledge and detail the key considerations which can ensure effective, efficient, and impactful implementation. This narrative review builds upon previous work (45, 46) by providing an overview of essential reasoning and justification that can help improve test selection, provide practical and scientific recommendations to ensure accurate and reproducible testing that can maximize the interpretation of physical qualities, and offer suggestions to promote optimal uptake of information. It will also provide examples and real-world evidence to support the interpretation of recommendations.

Selecting tests

Testing within strength and conditioning should be simple. Fundamentally, important physiological qualities should be assessed (e.g., speed or strength) and testing protocols should be completed consistently across time. While it may be tempting to try and make a test appear more ‘specific’ to a sport (e.g., adding a basketball free throw following a change of direction test), by altering a test, the assessment of the underlying physiological quality is often lost, and what is being quantified is no longer clear. Ironically, attempts to make a test more sport specific often undermine the development of an athlete because the test loses construct and ecological validity. Therefore, when testing athletes, the physiological quality must be identified (e.g., maximum strength or aerobic capacity) and practitioners should be comfortable knowing that a single test cannot assess all physiological capacities simultaneously.

The tests that coaches select and implement with their athletes should serve a purpose. Both athletes and coaches have limited time and the collection of data that is unusable or not maximized can be a waste of time and resources. Therefore, it is important to consider the test's purpose and what can be gleaned from its completion. To help guide practitioners in their selection of tests, it’s proposed that when assessing physical qualities, at least two (and ideally three) of the following concepts can be achieved:

1. Ranking
2. Monitoring
3. Prescription

These concepts, which are not listed in rank order of importance, help ensure that there is a purpose behind each assessment and that the test can be used to guide practice. See Figure 1 and the explanations below, which discuss ranking, monitoring, and prescription.

Insert Figure 1 Here

The ability to *rank* athletes is an important concept that helps guide athlete selection. *Ranking* refers to the concept that if two athletes from the same playing pool are compared, and all other physical qualities and technical/tactical skillsets are equal, the athlete with the greater ability in the tested quality should be ranked higher. It should be noted that the physical quality should be important for sporting performance or has established indirect relationships with performance. For example, a wide receiver in American football needs to have high levels of acceleration and maximum speed (63). Therefore, if two players were to be compared and all other physical qualities and technical/tactical skillsets were equal, the player with the greatest acceleration and maximum speed should be preferentially ranked as this would promote greater performance outcomes. Alternatively, if two rugby league players were to be compared, one who had high levels of lower body strength and another who had low levels of strength, it could be argued that the stronger athlete should be ranked higher than the weaker athlete. While rugby league is a complex sport, and the relationship between improvements in lower body strength and on-field performance is difficult to directly ascertain, greater strength is likely essential for helping mitigate the effects of collisions, support fundamental skills (e.g., wrestling within rucks), and support recovery post-match (18, 35, 82). Consequently, selecting tests that accurately measure fundamental and important qualities can be used to guide the ranking of athletes.

Grounded in the concept of reliability and sensitivity, selecting tests that allow practitioners to accurately monitor whether improvement has occurred is essential for longitudinal tracking. Ideally, the test should be reliable so that there are small amounts of noise (i.e., variability in performance) and sensitive enough to measure when an improvement in the physical quality has occurred. In tests that have a range of outcome measures (e.g., the countermovement jump), the use of highly variable metrics, such as rate of force development (RFD) (26), is not recommended as these make monitoring changes extremely difficult. It is acknowledged that theoretically an outcome variable can be interesting, but due to the variability associated with the measure, it is difficult to monitor.

Monitoring performance of a test should also be placed within the context of an athlete’s entire physical development. Tests can be confounded by a range of variables that, if not accounted for, may shroud the true change in an athlete’s performance. For example, athlete sprint times may not appear to improve over a collegiate career. However, when body mass is accounted for, it is clear that substantial improvements in momentum could have occurred (42). For collision sports, this is naturally a great advantage. Similarly, increases in body mass may mask improvements in aerobic capacity as athletes develop. However, increased body mass and maintenance in aerobic field tests can indicate greater running economy and improved high intensity running ability (11). Similar statements can be made for commonly implemented tests, such as the countermovement jump and corresponding kinetic variables (e.g., force), which can be strongly influenced by changes in body mass. Consequently, practitioners must carefully scrutinize their data beyond absolute values and understand the interaction of other physical qualities upon performance. This can not only provide an improved understanding of physical changes for practitioners but also reassure and educate athletes who have not seen the results they desire from a test.

Using testing information to guide training prescription should be a primary consideration for the strength and conditioning practitioner. The ability to test athletes, identify their strengths and weaknesses but then also improve their training is essential and tests have varying levels of application. For example, the 30-15 intermittent fitness test (30-15IFT) has greater application than a Yo-Yo intermittent recovery test (Yo-Yo IRT), as several programming tools have been developed and validated to guide prescription from this test (4). Alternatively, tests of maximal dynamic strength (e.g., 1-3RM back squat) have greater prescriptive utility than an isometric assessment (e.g., isometric mid-thigh pull (IMTP)), as strength coaches can prescribe loads as a percentage of maximum capacity using this information. Considering this, if faced with the need to assess a capacity, coaches should strategically select tests that allow for improved training prescription to help individualize and maximize the subsequent training block.

Validity, Reliability, and Sensitivity – The Heart of Athlete Testing

Fundamental to athlete testing and profiling are the concepts of validity, reliability, and sensitivity. If a test has low validity and/or reliability, the data collected are often a poor reflection of the individual's capacity or not a reflection of that capacity at all. Furthermore, when the sensitivity of a test is poor, interpretation of changes in the test between time points can be extremely difficult. Consequently, when considering whether to use a test, it is important to establish whether a test indeed measures that physical quality. Additionally, it is important to quantify the normal variation between assessments (i.e., the repeatability of the test and its outcomes).

Validity

Validity refers to whether a test indeed measures what it was designed to measure (32). There are several forms of validity which can be classified based on the accuracy of the outcome measure (i.e., test validity) or how trustworthy the protocols, conclusions, and generalizations are (i.e., methodological validity [often termed experimental validity in a research setting]; Figure 2). In Table 1, we detail the different forms of validity and how they relate to testing physical qualities.

*****Insert Figure 2 Here*****

*****Insert Table 1 Here*****

All types of validity are important when assessing an athlete's physical qualities and evidence for validity in several of its sub-domains is often necessary. With the growing uptake of sports technology for the monitoring of athletes, it is important to establish whether the equipment being used provides an accurate reflection compared to a standard measure (i.e., criterion validity) (86, 87). Recent reviews of global and local positioning systems (10) and commonly used resistance training monitoring devices (e.g., linear transducers, accelerometers) (90) have highlighted several concerns and considerations with these forms of technology. Specifically, these reviews highlight the importance of comparing devices to a ‘gold-standard’ criterion. This is important because if the criterion does not accurately reflect a measure, then the device that it is being compared to can have a misleading amount of error (either increased or decreased). Furthermore, it is essential to establish the accuracy of different outcome measures that are reported from technology. For example, when measuring back squat performance, mean and peak barbell velocity can both be assessed. However, a single device can report very different levels of accuracy dependent upon which outcome measure is used (9, 87, 91).

Threats to validity can occur not only from technology, but also the test instructions and protocols used. For example, when assessing accelerative ability with a 10 m sprint, starting an athlete 50 cm behind a timing gate or triggering timing using a front foot trigger (as is commonly done within practice and throughout the scientific literature (12, 92, 93)) substantially reduces the concurrent (criterion) validity, as these methods routinely miss ~20–50% of the athlete’s acceleration phase (89). In this instance, the criterion validity of the timing gates is not changed (i.e., the timing system is accurate), but modifications to the starting method have substantially altered the outcome. In a situation such as this, criterion validity becomes the victim but the issue stems from internal validity (i.e., the test design does not allow a true reflection of the observed results).

Conversely, on the opposite end of the methodological validity spectrum, ecological validity refers to how well a test relates to actual athlete performance and whether it can be applied to real-life settings. For instance, asking field hockey athletes to complete a cycling time trial to establish VO_{2max} has limited ecological validity. Alternatively, a field-based running assessment (e.g., 30-15 IFT) may be more appropriate. This example additionally highlights the consideration for construct validity. Coaches may use tests such as lab-based VO_{2max} assessment or the 30-15 IFT to assess cardiovascular “fitness”. The former achieves this via direct measurement of aerobic capacity, while the latter is a construct within itself (high-intensity intermittent running ability) that is comprised of aerobic capacity, as well as other physical qualities such as anaerobic and neuromuscular qualities. Therefore, it is important for practitioners to understand which physical constructs are being assessed and the extent to which the tests used are an accurate representation of the definitions of that construct.

Reliability

Reliability refers to the degree of repeatability, reproducibility, or consistency in a measure (49, 102). A test outcome can be reliable even if it is not valid (Figure 3), but if it is not reliable then it cannot be valid. To be able to assess changes in performance, the reliability of the test needs to be established (test-retest reliability). If a test cannot be reliably reproduced, coaches cannot confidently state whether an athlete has truly improved in a test.

Insert Figure 3 Here

As with internal validity, a range of factors can influence reliability, and these factors are often unique to a test or a specific outcome measure. For example, jump height during the countermovement jump could be influenced by the instructions provided to the athlete (37, 62), the method of calculation (e.g., flight time *vs.* impulse-momentum relationship *vs.* take-off velocity) (55), or the technology used (60). Alternatively, for anthropometry and body composition, food or fluid consumption could alter outcomes and should be standardized across days (59). Consequently, to make accurate inferences about changes in performance, coaches should quantify the reliability of each test and outcome measure with their cohort of athletes or have strong grounds to justify the reliability from a similar cohort in the literature (8, 65, 68). Recommendations for enhancing test reliability and reducing measurement error are supplied in Table 2.

Insert Table 2 Here

For tests of physical performance or capacity, it is recommended that the reliability of a test is established across the time period that data will be routinely collected and interpreted (i.e., between-day reliability). Typically, longer periods between test-retest assessments result in less reliable outcomes. This has implications for tests of a more exhaustive nature, such as those assessing maximal high-intensity intermittent running ability, which are typically performed >6 weeks apart (50). Furthermore, it is important to test in a standardized state (e.g., 48 hours of rest prior to the test) and when changes in physical performance/capacity would not be thought to have changed (e.g., following strenuous exercise). If human error can be introduced through assessment (e.g., skinfold measurements for estimates of body composition), intra-rater and inter-rater reliability should be quantified and, if possible, minimized. To reduce this variability and improve measurement reliability, all assessors should be adequately trained (e.g., International Society for the Advancement of Kinanthropometry (ISAK) for body composition measurements), and changes in assessor between pre- and post-measurements should be avoided if possible. Finally, environmental conditions (e.g., temperature, wind, testing surface) should be standardized to enhance the reliability of physical performance testing. Naturally, this can be difficult when testing outdoors. Therefore practitioners should carefully consider where and when testing occurs.

One final consideration of testing that is often reported but poorly disseminated within the literature is the reliability of technology. Considering that technology is commonly used during testing, there is a need to establish the between-device error and between-day error. However, to accurately reflect the error of the technology being assessed, it is essential not to attribute biological error to technological error (86, 90). For example, if a practitioner wishes to establish the reliability of a linear position transducer for the measurement of mean concentric barbell velocity during the back squat, it is important to delineate between the variability of exercise performance during the squat and the error of the technology. This minimizes the risk of inappropriate attribution of error to technology when it may just be that humans struggle to replicate a task perfectly (i.e., normal performance variation). Recent reviews (10, 90) have emphasized this point and strongly recommended that to measure the reliability of a measurement device appropriately, human error should be eliminated.

The reliability of a test outcome measure can be quantified with various statistics, such as the standard error of measurement (SEM; sometimes referred to as the typical error) or the intraclass correlation coefficient (ICC). While both these statistics are recommended to paint the full picture of reliability (i.e., absolute vs. relative reliability, respectively), the SEM is perhaps more useful in practice as it provides an estimate of the within-athlete variability (i.e., how much athletes typically fluctuate by in their test performance over the retest period). Approximate between-day coefficient of variations (CV; the SEM expressed as a percentage of the mean) for commonly used physical capacity tests are provided in Table 3. The CV or SEM can be used to assess test sensitivity, such as tracking changes within an individual. Full details on reliability analysis and applications can be found elsewhere (30, 49).

Insert Table 3 Here

Sensitivity

Test sensitivity, or responsiveness, refers to the ability of a test to detect real and important changes in performance. It is implicitly linked to both validity and reliability for several reasons. First, if a test does not possess adequate test or methodological validity, then changes in the outcome measure may occur despite no real changes in an athlete’s physiology or performance capacity. Second, when the outcome measure of a test is a construct itself, it may be difficult to identify changes in specific, underlying physiological qualities. For instance, the 30-15 IFT is commonly used as a marker of cardiorespiratory or aerobic fitness in team sports. But since the test assesses maximal intermittent high-intensity running ability, final running performance (V_{IFT}) is also determined by anaerobic and neuromuscular qualities. Therefore, tests in which the outcome measure is a construct may lack sensitivity to isolated physiological systems. Despite this, such tests may still be considered useful.

The reliability of a test outcome also has implications for responsiveness. Reliability determines the noise of a test, which is needed to help understand whether the changes in performance are ‘real’ or simply the result of test error/biological variation. In addition, the ‘smallest worthwhile change’ (SWC) must be established. Thresholds for a worthwhile change are primarily established through two methods. These are:

1) Anchor-based
2) Distribution-based

Ideally, an anchor-based method should be implemented, as it holds high levels of ecological validity while also allowing practitioners to relate changes in training and testing data to real-world outcomes. Anchor-based SWC can be established through prognostic- or validity studies in which a measure has been used to predict an outcome and can be found within the literature (79) or through an opinion-based method in which an expert (e.g., an established practitioner) in the field provides an estimate of what would be deemed a meaningful, real-world change (14, 40). These thresholds are often also named the minimum practical or clinically important difference.

When an anchor-based approach is not feasible, a distribution-based method could be implemented. This method quantifies the typical deviation in how athletes perform between each other (i.e., the between-athlete standard deviation [SD]) and a fraction of this is used to represent the change required to meaningfully alter their position within this distribution. Commonly, 0.2 · between-athlete SD is calculated to detect the SWC. Further, 0.6 and 1.2 are often used for moderate and large changes, respectively (28). This method is commonplace in the sport science literature, perhaps due to the lack of studies or quality evidence for anchor-based approaches. However, we warn practitioners and coaches that a 'blanket' target change of 0.2 of the between-athlete SD systematically underestimates practically relevant and more informed changes from the methods previously described (14). We therefore advocate anchor-based approaches, which can be informed by literature, empirical research findings (in-house or published), and internal discussions between the entire performance, coaching and medical team.

Once a meaningful change has been calculated/established, consideration of the reliability of the test relative to the observed change can occur. Tests that have high levels of reliability have greater likelihood of being able to infer a 'real' change. First, the error of the test can simply be scaled in relation to the SWC to determine its 'usefulness'. The Australian Institute of Sport (AIS) has historically rated tests as:

- 'Good' – When the SEM of the test is less than the smallest meaningful change.
- 'OK' – When the SEM of the test is approximately the same as the smallest meaningful change.

- ‘Marginal’ – When the SEM of the test is much greater than the smallest meaningful change

There are several other ways in which practitioners can determine the certainty of a change. Perhaps the most informative is visualizing the test change against both its error (noise) and the SWC. In this process, there are a few simple but effective methods that can help inform the interpretation of a test:

1. An observed test outcome can be visualized with its SEM derived from a test-retest reliability study or similar. The SEM represents within-athlete variability under ‘normal’ or ‘standardized’ conditions (Figure 4).
2. A change in the test outcome between two measurement points can be visualized with the adjusted SEM, which is the usual SEM multiplied by the square root of two. This correction accounts for the fact that the change score must incorporate error from both testing occasions (test one and test two). Naturally, this makes the adjusted SEM larger (~1.4 times) than the observed SEM (Figure 5).
3. The adjusted SEM for a change can be converted into compatibility limits (CL), which provide a range of values compatible with the test error. There are many resources available describing how this process can be achieved (29, 49, 88). The CL can be specified at a given ‘level’ that defines the coverage probability (i.e., how much of the distribution is covered). For example, a 100% CL would cover all the distribution, whereas the SEM alone is equivalent to only a 68% CL. There is no right or wrong answer as to which CL is optimum and it depends on how conservative a practitioner wishes to be when interpreting the data. Our recommendation would be values between 80–90% (Figure 6).

*** Insert Figures 4-6 Here***

Order of testing

The order of a testing battery can substantially alter the validity of the testing outcomes. For example, if a highly fatiguing test (e.g., a maximal aerobic test) is completed before another test (e.g., a sprint), the second test's performance will likely suffer. Naturally, this can have ramifications for identifying performance changes; consequently, the standardization of testing order is essential for the accurate, reproducible, and fair assessment of physical performance.

The order of tests should be determined by the physiological demands placed on the athlete. Completing one test should have minimal impact on the performance of subsequent tests, with tests that require minimal recovery (e.g., anthropometry, short efforts) placed before more physically demanding tests (102). Furthermore, for the sake of feasibility and efficiency, there also needs to be an appropriate 'flow' within the testing order. Or in other words, athletes should not be required to undergo substantial logistics to undertake testing.

Naturally, the tests that each sport and athlete require differs. Furthermore, the available time can also alter the number of tests completed. Thus, the physical qualities that have the largest contribution or influence on athletic performance should be prioritized. However, across nearly all sports, the assessment of fundamental physical qualities (e.g., strength, power, aerobic and anaerobic capacity) is valuable. Consequently, Figure 7 provides recommendations on the order of tests considering these fundamental qualities (45).

Insert Figure 7 Here

Maximizing the outcomes from testing

Practitioners often have limited time to test athletes. While testing is an important step in physical development, due to the many requirements that athletes face, windows of opportunity are often limited. Therefore, there is a need to maximize the amount of information that can be attained from a small number of tests that can help the ranking of athletes, the monitoring of physical characteristics, and the prescription of training. Maximizing testing data can be achieved through a range of methods, including the strategic selection of tests, the outcome measures recorded, and the equipment used. While technology should not be used just because it is available, if the technology enables greater insight into an athlete's physical qualities when they perform the same test, practitioners should consider its use. Furthermore, by carefully considering how and what tests are being implemented, practitioners can have a substantial improvement in the efficiency of testing while improving the impact for coaches.

The inclusion of certain forms of technology can help improve the information that can be attained from testing, with little to no additional effort from the athletes involved. An obvious example is the inclusion of a force plate over a Vertec to assess jump performance so that additional important kinetic and kinematic information can be quantified. However, other technology includes using laser/radar devices, linear position transducers, mobile applications, and global positioning systems to enhance testing outcomes. For example, if linear sprint testing is already occurring, the addition of laser/radar technology that can measure athlete instantaneous time-displacement data can provide a wealth of information regarding an athlete's horizontal force-velocity-power profiles (57). Moreover, this information can be used to identify deficiencies in physical capacity and justify whether greater high force (e.g., heavy sled pull/pushing) or high velocity (e.g., unresisted maximal sprints) exercises are required (27). Alternatively, if a laser is not available, but a team uses GPS, an athlete's peak velocity can be attained to guide decisions around exposure to sprinting during training or paired with an athlete's maximal aerobic speed to provide their anaerobic speed reserve (67). Finally, during resistance training, if athletes are already completing maximal strength testing (e.g., 1 repetition maximum (1RM) in the bench press or squat), the inclusion of a device that can accurately measure barbell velocities during the submaximal loads (e.g., 25, 50, 75% of 1RM) can support the development of a load-velocity profile (2). This information can be used to regulate resistance training loads and volumes better and help mitigate the risk of training to failure. Furthermore, it can support monitoring changes in strength/power characteristics across time.

A simple method of enhancing the recorded data can be through 'pairing' outcomes from tests together so that data can be used to infer additional information. For example, by calculating mean sprint velocity from the times retrieved during linear sprint testing, then multiplying this value with body mass, initial and peak sprint momentum can be calculated. This information is a valid discriminator between professional and sub-professional athletes and may be useful for monitoring long-term changes in physical capacity (36, 42). Alternatively, the consideration of body mass during tests of aerobic capacity, such as the 30-15 IFT, may help account for the influence of body mass and demonstrate to an athlete that there has been an improvement in high intensity running performance despite not necessarily attaining a higher score (11). Outside the addition of body mass, simple strength, and power measures can be combined to help quantify performance and guide training. For example, the dynamic strength index can be calculated by comparing peak force from the IMTP and the countermovement jump/squat jump, and may be useful in justifying whether additional strength or plyometric work would be beneficial (7). Alternatively, the eccentric utilization ratio, which uses the performance from eccentric-concentric and concentric-only exercises (e.g., countermovement jump and the squat jump) in a ratio, could be useful in guiding practitioners in whether athletes are effectively using the eccentric portion of a movement (47). Although, it should be noted that the dynamic strength index and eccentric utilization ratio should be contextualized, with each sub-component scrutinized (74, 75). Consequently, practitioners and researchers should carefully consider whether the strategic combination of data can enhance testing outcomes. Table 4 provides information regarding technology and measures that can be easily used to attain additional testing outcomes.

Insert Table 4 Here

'Invisible monitoring' and its use in testing

The concept of 'invisible monitoring' (i.e., testing athletes as they train and perform without specific intervention) has had substantial interest in recent times (20, 41, 71, 97). Organizing and coordinating testing opportunities with coaches, players, and support staff can be time-consuming and stressful. Therefore, understanding an athlete's physical capacity without intervening is highly valued. The use of wearable microtechnology and monitoring equipment has allowed continual, non-invasive assessment of qualities without having to make extensive alterations to training. By using technology to invisibly monitor performance during exercise, practitioners have more regular information regarding their athletes and can also use this information to detect changes across time. They can also make better-informed decisions if prior testing data is poor/inaccurate (e.g., if an athlete is demotivated or performance during a testing occasion simply does not reflect their true capacity). The ability to test/monitor physical changes can occur during warmups or the main training session, depending on what is being monitored (e.g., changes in strength, aerobic adaptations).

Identifying opportunities to monitor changes in important physical qualities is integral to invisible monitoring. For example, peak velocity can be assessed during training through the use of GPS (64). If speed is an important quality for a given sport, practitioners often expose athletes to maximal sprinting efforts during training to develop this quality. Therefore, coaches may wish to include a maximal effort at the end of a warmup or the start of a training session and monitor changes in peak velocity across time using GPS data (64). By doing this, the coaches gain important information around the development of this quality. Furthermore, if changes occur, this data can guide decisions around relative exercise intensity and subsequent training prescription (e.g., anaerobic speed reserves). On the other hand, the use of submaximal fitness tests have been proposed as a feasible alternative to maximal fitness tests to evaluate an athlete’s physiological state. While the reader is directed towards the review by Shushan and colleagues (71) for a thorough explanation of their implementation, submaximal fitness tests have the potential to be administered to a group of athletes as part of a warmup to help detect changes in cardiorespiratory and endurance performance. These tests are far less intensive than traditional methods of assessing endurance performance (e.g., Yo-Yo intermittent recovery test) and can be completed in as little as 3-4 minutes with standardized distances and velocities used to help reduce setup time (71).

During resistance training, changes in strength and power can regularly be assessed through monitoring the kinetic and kinematic outputs produced with submaximal loads at the end of a warmup or throughout a training session. Due to the linear and relatively stable load-velocity relationship and the knowledge that velocity at 1RM shows minimal variation within- and between-athletes (2, 19, 33, 66), changes in the velocities with submaximal loads can infer improvements in maximal strength/power qualities. Examples include monitoring the changes in barbell velocity with a set load (e.g., 100kg) at the end of a warmup, measuring changes in set loads based on a previously constructed load-velocity profile, or using multiple loads and velocities from a warmup to estimate changes in maximal strength (e.g., implementation of the ‘2-point method’) (3, 21, 88, 98). These methods can all be done outside of usual testing and can be implemented with little to no alteration to training. Furthermore, they offer viable and pragmatic solutions beyond setting aside specific testing occasions to help practitioners gain regular updates on their athlete’s physical qualities. An example of testing data from an athlete’s warmup is compared to data recorded from an original testing occasion in Figure 8, with this data suggesting that changes in their strength have occurred.

Insert Figure 8 Here

Presentation of data

The presentation of testing data is both a science and an art. There is science behind how humans process, analyze, and subsequently interpret data (69, 83). But the presentation and actual visualization of data is an art in which you can present information, communicate an idea, and persuade the viewer if needed. This is particularly pertinent within sports science, as effectively designed data visualizations allow the viewer (often a coach or athlete) to quickly understand key points and patterns across large swathes of data (5). However, ineffectively designed visualizations can cause misunderstanding and, potentially, distrust.

Therefore, the presentation of testing information can be just as important as the testing itself. While a range of methods can be used to enhance testing data, this section provides recommendations to help improve data presentation so that athletes and practitioners can understand testing outcomes.

When presenting information, the most important considerations are *who is the audience* and *what is the purpose* of presenting this data. ‘Understanding your audience’ includes 1) knowing their preference of data presentation (e.g., do they want a quick visual or do they want to know every single number?) and 2) establishing what level of understanding they have of this type of information (e.g., does a head coach know what the test is trying to measure and why it matters to performance?). Furthermore, establishing *why* the data is being presented can ensure the information is clear and differences, or the lack of them, can be emphasized. Consequently, considering *who* the audience is and the purpose of presenting the information, you can best guide the viewer to reach the right conclusion and help influence decision-making.

The presentation of data should be as simple, effective, and efficient as possible. Time is often the biggest constraint in sport; therefore, keeping testing data simple and informative so that maximal information is quickly conveyed is advantageous. Considering this, a range of methods (refer to Table 5) can be used to emphasize certain points and help convey a message. Furthermore, visual processing of data can be substantially enhanced when several of these methods are combined strategically (83). For example, information can be portrayed more efficiently, and the cognitive load can be reduced when the information is colored, grouped, and enclosed to show discrete differences (refer to Figure 9). Alternatively, data (e.g., player performance) could be more easily interpreted when colour, size, and grouping of the data are combined (refer to Figure 10). Moreover, to reduce the cognitive load on the viewer, visual presentations should emphasize the key points, while surplus information that is not integral should be minimized or removed. Classic examples of ‘figure clutter’ that can impede the processing of information includes gridlines, tick marks, unnecessary data labels, and 3D effects (69). While simple edits, such as the rotation of axis titles and the provision of specific information relating to performance that would not be easily ascertained (e.g., the use of the specific velocities attained in Figure 10F) helps to remove any uncertainty in performance.

Insert Figure 9 Here

Insert Figure 10 Here

Insert Table 5 Here

To help accentuate the value of the data being presented, providing as much context as pragmatically possible can help coaches to understand the meaning of the data. Even experienced practitioners and researchers will have a poor understanding of a single value (or set of values) when performance is not placed into context. This may include information that compares the performance of players of a similar position, playing level, or the wider population. Additionally, graphical illustrations that emphasize the magnitude of differences in certain physical qualities between athletes can be valuable. For example, a difference of 0.2 seconds could sound trivial to many coaches. However, others will know that 0.2 seconds in a 20 m sprint is a large improvement/difference. Alternatively, a $0.2 m \cdot s^{-1}$ mean concentric barbell velocity difference in the back squat may sound small, but in reality, it suggests a difference of ~15-20% 1RM (22). A range of statistical methods are available to help illustrate these differences (e.g., Z-scores and T-Scores (46, 84); refer to Figures 11 and 12) and demonstrating the magnitude of difference, irrespective of the units of measurement can effectively illustrate the practical significance of the data being presented.

Insert Figure 11

Insert Figure 12

Finally, perhaps of greatest importance to the maximization of the collected data is the speed with which the information can be returned to those who require it. It is well established that immediate augmented feedback during exercise can support the execution and improvement of physical performance (94, 95, 100, 101); and the provision of testing data to coaches is no different. Time delays in the provision of feedback mitigates its usefulness, with the usefulness of information inversely related to the turnaround time between the performance and when it is available to the user (Figure 13) (31). Consequently, it is prudent for sport scientists and strength and conditioning coaches to clearly establish when data will be returned, with information from testing ideally being made available as soon as feasibly possible so that coaches can make informed decisions around training programs and prescription. The longer the delay in returning testing information, the less useful that testing occasion is.

Insert Figure 13 Here

"It's important, but it's not everything" – Understanding the importance and role of testing

Undeniably, well-developed physical qualities are important and often essential for high-level performance. However, it's also crucial to acknowledge that they are only one aspect of sporting success (44). While athletic development and performance in tests of physical qualities can be incredibly alluring, they do not always transfer to improved outcomes. Indeed, it should be acknowledged that performance on physical testing batteries often only makes up a portion of the selection picture, with sport-specific skill extremely important. Consequently, strength and conditioning coaches should understand the perceptions of fitness testing and physical qualities and how it fits within the holistic development of the athlete. Therefore, while appropriately selected testing batteries should be used to guide selection decisions, monitor changes in physical qualities, and support training prescription, chasing numbers for the sake of improvement on a test or setting arbitrary thresholds/standards for players to attain, may be counterproductive. Instead, it is recommended that strength and conditioning coaches work alongside a multi-disciplinary team and use testing results to guide decisions and drive conversations within context rather than letting the results dictate them.

Conclusions

The testing of physical qualities is fundamental to strength and conditioning and can help improve the chances of success for an athlete or team. Information from testing can support coaches in their selection of athletes, the prescription of training, and the assessment of whether training interventions are working. However, it is essential the tests that are being implemented are selected for the right reasons. Fundamental concepts such as validity, reliability, and sensitivity need to be well understood so that decisions are made from accurate and reproducible testing information. Furthermore, understanding *why* testing is being completed, *how* the testing is being executed, and *what* outcomes will occur from this information can substantially improve the odds of implementing a successful testing battery.

When well-designed testing batteries are employed, a host of previously unavailable information becomes accessible. Strategic selection of outcome measures, use of technology, and awareness from coaching staff can help maximize information about athletes and help to provide regular updates about physical qualities. Additionally, through good data handling practices and clever presentation, testing information can be efficiently portrayed to athletes and colleagues to convey important points and help influence decisions around physical development. While it is acknowledged that testing of athletes can be stressful, the decisions around the tests used and actual outcome measures retrieved should be simple. To help guide these decisions, Figure 14 provides a simple flowchart to help coaches decide whether the test should be implemented.

Insert Figure 14 Here

References

1. Bangsbo J, Iaia FM, and Krustrup P. The Yo-Yo intermittent recovery test: a useful tool for evaluation of physical performance in intermittent sports. *Sports Med* 38: 37-51, 2008.
2. Banyard HG, Nosaka K, and Haff GG. Reliability and Validity of the Load-Velocity Relationship to Predict the 1RM Back Squat. *J Strength Cond Res* 31: 1897-1904, 2017.
3. Banyard HG, Tufano JJ, Weakley JJS, et al. Superior Changes in Jump, Sprint, and Change-of-Direction Performance but Not Maximal Strength Following 6 Weeks of

Velocity-Based Training Compared With 1-Repetition-Maximum Percentage-Based Training. *Int J Sports Physiol Perform* 16: 232-242, 2020.

4. Buchheit M. The 30-15 intermittent fitness test: accuracy for individualizing interval training of young intermittent sport players. *J Strength Cond Res* 22: 365-374, 2008.
5. Buchheit M. Want to see my report, coach? *Aspetar Sport Med J* 6: 34-43, 2017.
6. Byrne DJ, Browne DT, Byrne PJ, and Richardson N. Interday Reliability of the Reactive Strength Index and Optimal Drop Height. *J Strength Cond Res* 31: 721-726, 2017.
7. Comfort P, Thomas C, Dos'Santos T, et al. Changes in Dynamic Strength Index in Response to Strength Training. *Sports* 6: 176, 2018.
8. Cormack SJ, Newton RU, McGuigan MR, and Doyle TL. Reliability of measures obtained during single and repeated countermovement jumps. *Int J Sports Physiol Perf* 3: 131-144, 2008.
9. Crang ZL, Duthie G, Cole MH, et al. The inter-device reliability of global navigation satellite systems during team sport movement across multiple days. *J Sci Med Sport*, 2021.
10. Crang ZL, Duthie G, Cole MH, et al. The Validity and Reliability of Wearable Microtechnology for Intermittent Team Sports: A Systematic Review. *Sports Med* 51: 549-565, 2021.
11. Darrall-Jones J, Roe G, Carney S, et al. The Effect of Body Mass on the 30-15 Intermittent Fitness Test in Rugby Union Players. *Int J Sports Physiol Perform* 11: 400-403, 2016.
12. Darrall-Jones JD, Jones B, Roe G, and Till K. Reliability and Usefulness of Linear Sprint Testing in Adolescent Rugby Union and League Players. *J Strength Cond Res* 30: 1359-1364, 2016.
13. Darrall-Jones JD, Jones B, and Till K. Anthropometric and Physical Profiles of English Academy Rugby Union Players. *J Strength Cond Res* 29: 2086-2096, 2015.
14. Datson N, Lolli L, Drust B, et al. Inter-methodological quantification of the target change for performance test outcomes relevant to elite female soccer players. *Sci Med Football* 6: 248-261, 2022.
15. Deprez D, Coutts AJ, Lenoir M, et al. Reliability and validity of the Yo-Yo intermittent recovery test level 1 in young soccer players. *J Sports Sci* 32: 903-910, 2014.
16. Edwards T, Weakley J, Woods CT, et al. Comparison of Countermovement Jump and Squat Jump Performance Between 627 State and Non-State Representative Junior Australian Football Players. *J Strength Cond Res*: 10.1519, 2022.
17. Franconeri SL, Padilla LM, Shah P, Zacks JM, and Hullman J. The science of visual data communication: What works. *Psyc Sci Pub Int* 22: 110-161, 2021.
18. Gabbett TJ and Seibold AJ. Relationship between tests of physical qualities, team selection, and physical match performance in semiprofessional rugby league players. *J Strength Cond Res* 27: 3259-3265, 2013.
19. Garcia-Ramos A, Barboza-Gonzalez P, Ulloa-Diaz D, et al. Reliability and validity of different methods of estimating the one-repetition maximum during the free-weight prone bench pull exercise. *J Sports Sci* 37: 2205-2212, 2019.
20. García-Ramos A, Jukic I, Weakley J, and Janićijević D. Bench press one-repetition maximum estimation through the individualised load-velocity relationship: comparison of different regression models and minimal velocity thresholds. *Int J Sports Physiol Perf* 16, 2021.
21. García-Ramos A, Weakley J, Janicijevic D, and Jukic I. Number of Repetitions Performed Before and After Reaching Velocity Loss Thresholds: First Repetition

Versus Fastest Repetition-Mean Velocity Versus Peak Velocity. *Int J Sports Physiol Perform* 16: 950-957, 2021.

22. González-Badillo JJ and Sánchez-Medina L. Movement velocity as a measure of loading intensity in resistance training. *Int J Sports Med* 31: 347-352, 2010.
23. Harrison PW and Johnston RD. Relationship Between Training Load, Fitness, and Injury Over an Australian Rules Football Preseason. *J Strength Cond Res* 31: 2686-2693, 2017.
24. Haugen T and Buchheit M. Sprint Running Performance Monitoring: Methodological and Practical Considerations. *Sports Med* 46: 641-656, 2016.
25. Haugen T, Tønnessen E, Hisdal J, and Seiler S. The role and development of sprinting speed in soccer. *Int J Sports Physiol Perform* 9: 432-441, 2014.
26. Heishman AD, Daub BD, Miller RM, et al. Countermovement jump reliability performed with and without an arm swing in NCAA division 1 intercollegiate basketball players. *J Strength Cond Res* 34: 546-558, 2020.
27. Hicks DS, Schuster JG, Samozino P, and Morin J-B. Improving mechanical effectiveness during sprint acceleration: practical recommendations and guidelines. *Strength Cond J* 42: 45-62, 2020.
28. Hopkins W, Marshall S, Batterham A, and Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Ex* 41: 3, 2009.
29. Hopkins WG. Individual responses made easy. *J App Physiol* 118: 1444-1446, 2015.
30. Hopkins WG. Spreadsheets for Analysis of Validity and Reliability. *Sportscience* 19: 36-42, 2017.
31. Hubbard M. Computer simulation in sport and industry. *J Biomech* 26: 53-61, 1993.
32. Impellizzeri FM and Marcora SM. Test validation in sport physiology: lessons learned from clinimetrics. *Int J Sports Physiol Perf* 4: 269-277, 2009.
33. Janicijevic D, Jukic I, Weakley J, and García-Ramos A. Bench Press 1-Repetition Maximum Estimation Through the Individualized Load–Velocity Relationship: Comparison of Different Regression Models and Minimal Velocity Thresholds. *Int J Sports Physiol Perf* 16: 1074-1081, 2021.
34. Jaric S and Garcia Ramos A. Letter to the editor concerning the article "Bar velocities capable of optimising the muscle power in strength-power exercises" by Loturco, Pereira, Abad, Tabares, Moraes, Kobal, Kitamura & Nakamura (2017). *J Sports Sci* 36: 994-996, 2018.
35. Johnston RD, Gabbett TJ, Jenkins DG, and Hulin BT. Influence of physical qualities on post-match fatigue in rugby league players. *J Sci Med Sport* 18: 209-213, 2015.
36. Jones B, Weaving D, Tee J, et al. Bigger, stronger, faster, fitter: the differences in physical qualities of school and academy rugby union players. *J Sports Sci* 36: 2399-2404, 2018.
37. Kershner AL, Fry AC, and Cabarkapa D. Effect of internal vs. external focus of attention instructions on countermovement jump variables in NCAA Division I student-athletes. *J Strength Cond Res* 33: 1467-1473, 2019.
38. Krustrup P, Mohr M, Amstrup T, et al. The yo-yo intermittent recovery test: physiological response, reliability, and validity. *Med Sci Sports Ex* 35: 697-705, 2003.
39. Krustrup P, Mohr M, Nybo L, et al. The Yo-Yo IR2 test: physiological response, reliability, and application to elite soccer. *Med Sci Sports Ex* 38: 1666-1673, 2006.
40. Kyprianou E, Lolli L, Haddad HA, et al. A novel approach to assessing validity in sports performance research: integrating expert practitioner opinion into the statistical analysis. *Sci Med Football* 3: 333-338, 2019.

41. Leduc C, Tee J, Lacome M, et al. Convergent Validity, Reliability, and Sensitivity of a Running Test to Monitor Neuromuscular Fatigue. *Int J Sports Physiol Perform* 15: 1067-1073, 2020.
42. Mann JB, Mayhew JL, Dos Santos ML, Dawes JJ, and Signorile JF. Momentum, Rather Than Velocity, Is a More Effective Measure of Improvements in Division IA Football Player Performance. *J Strength Cond Res* 36: 551-557, 2022.
43. Markovic G, Dizdar D, Jukic I, and Cardinale M. Reliability and factorial validity of squat and countermovement jump tests. *J Strength Cond Res* 18: 551-555, 2004.
44. McCormack S, Jones B, Scantlebury S, Rotheram D, and Till K. "It's Important, but It's Not Everything": Practitioners' Use, Analysis and Perceptions of Fitness Testing in Academy Rugby League. *Sports* 8: 130-148, 2020.
45. McGuigan M. Principles of Test Selection and Administration. in: *Essentials of Strength Training and Conditioning*. GG Haff, NT Triplett, eds. Champaigne, IL, USA: Human Kinetics, 249-258, 2015.
46. McGuigan MR, Cormack SJ, and Gill ND. Strength and power profiling of athletes: Selecting tests and how to use the information for program design. *Strength Cond J* 35: 7-14, 2013.
47. McGuigan MR, Doyle TL, Newton M, et al. Eccentric utilization ratio: effect of sport and phase of training. *J Strength Cond Res* 20: 992-995, 2006.
48. McLaren S, Coutts AJ, and Impellizzeri FM. Perception of Effort and Subjective Monitoring. in: *NSCA's Essentials of Sport Science*. DN French, L Torres-Ronda, eds. Champaign, IL: Human Kinetics, 231–254, 2021.
49. McLaren SJ. Reliability and Measurement Error. in: *Sport and Exercise Physiology Testing Guidelines: Volume II*. R Davison, P Smith, eds. London: Routledge, 48-52, 2022.
50. McLaren SJ, Smith A, Bartlett JD, Spears IR, and Weston M. Differential training loads and individual fitness responses to pre-season in professional rugby union players. *J Sports Sci* 36: 2438-2446, 2018.
51. McMahon JJ, Jones PA, and Comfort P. Comment on:"Anthropometric and physical qualities of elite male youth rugby league players". *Sports Med* 47: 2667-2668, 2017.
52. McMahon JJ, Murphy S, Rej SJ, and Comfort P. Countermovement jump phase characteristics of senior and academy rugby league players. *Int J Sports Physiol Perf* 12: 803-811, 2017.
53. Merrigan JJ, Stone JD, Hornsby WG, and Hagen JA. Identifying reliable and relatable force–time metrics in athletes—considerations for the isometric mid-thigh pull and countermovement jump. *Sports* 9: 4-17, 2020.
54. Midway SR. Principles of Effective Data Visualization. *Patterns* 1: 100141, 2020.
55. Moir GL. Three different methods of calculating vertical jump height from force platform data in men and women. *Meas Phys Ed Ex Sci* 12: 207-218, 2008.
56. Moore DA, Jones B, Weakley J, Whitehead S, and Till K. The field and resistance training loads of academy rugby league players during a pre-season: Comparisons across playing positions. *Plos one* 17: e0272817, 2022.
57. Morin J-B and Samozino P. Interpreting power-force-velocity profiles for individualized and specific training. *Int J Sports Physiol Perf* 11: 267-272, 2016.
58. Morrison M, Martin DT, Talpey S, et al. A systematic review on fitness testing in adult male basketball players: tests adopted, characteristics reported and recommendations for practice. *Sports Med* 52: 1491–1532, 2022.
59. Nana A, Slater GJ, Hopkins WG, et al. Importance of standardized DXA protocol for assessing physique changes in athletes. *Int J Sport Nut Ex Metab* 26: 259-267, 2016.

60. Nuzzo JL, Anning JH, and Scharfenberg JM. The reliability of three devices used for measuring vertical jump height. *J Strength Cond Res* 25: 2580-2590, 2011.
61. Pérez-Castilla A, Rojas FJ, and García-Ramos A. Assessment of unloaded and loaded squat jump performance with a force platform: Which jump starting threshold provides more reliable outcomes? *J Biomech* 92: 19-28, 2019.
62. Pérez-Castilla A, Weakley J, García-Pinillos F, Rojas FJ, and García-Ramos A. Influence of countermovement depth on the countermovement jump-derived reactive strength index modified. *Eur J Sport Sci* 21: 1606-1616, 2021.
63. Robbins DW. Positional physical characteristics of players drafted into the National Football League. *J Strength Cond Res* 25: 2661-2667, 2011.
64. Roe G, Darrall-Jones J, Black C, et al. Validity of 10-HZ GPS and Timing Gates for Assessing Maximum Velocity in Professional Rugby Union Players. *Int J Sports Physiol Perform* 12: 836-839, 2017.
65. Roe G, Darrall-Jones J, Till K, et al. Between-days reliability and sensitivity of common fatigue measures in rugby players. *Int J Sports Physiol Perf* 11: 581-586, 2016.
66. Sanchez-Medina L and González-Badillo JJ. Velocity loss as an indicator of neuromuscular fatigue during resistance training. *Med Sci Sports Ex* 43: 1725-1734, 2011.
67. Sandford GN, Kilding AE, Ross A, and Laursen PB. Maximal sprint speed and the anaerobic speed reserve domain: the untapped tools that differentiate the world's best male 800 m runners. *Sports Med* 49: 843-852, 2019.
68. Sawczuk T, Jones B, Scantlebury S, et al. Between-day reliability and usefulness of a fitness testing battery in youth sport athletes: Reference data for practitioners. *Meas Phys Ed Ex Sci* 22: 11-18, 2018.
69. Schwabish J. The Practice of Visual Data Communication: What Works. *Psyc Sci Pub Int* 22: 97-109, 2021.
70. Scott TJ, Duthie GM, Delaney JA, et al. The Validity and Contributing Physiological Factors to 30-15 Intermittent Fitness Test Performance in Rugby League. *J Strength Cond Res* 31: 2409-2416, 2017.
71. Shushan T, McLaren SJ, Buchheit M, et al. Submaximal Fitness Tests in Team Sports: A Theoretical Framework for Evaluating Physiological State. *Sports Med*, 2022.
72. Smart DJ, Hopkins WG, and Gill ND. Differences and changes in the physical characteristics of professional and amateur rugby union players. *J Strength Cond Res* 27: 3033-3044, 2013.
73. Stone MH, Sands WA, Pierce KC, et al. Relationship of maximum strength to weightlifting performance. *Med Sci Sports Exerc* 37: 1037-1043, 2005.
74. Suchomel TJ, Sole CJ, Bellon CR, and Stone MH. Dynamic strength index: Relationships with common performance variables and contextualization of training recommendations. *J Hum Kinet* 74: 59-70, 2020.
75. Suchomel TJ, Sole CJ, and Stone MH. Comparison of Methods That Assess Lower-body Stretch-Shortening Cycle Utilization. *J Strength Cond Res* 30, 2016.
76. Thomas A, Dawson B, and Goodman C. The yo-yo test: reliability and association with a 20-m shuttle run and VO2max. *Int J Sports Physiol Perf* 1: 137-149, 2006.
77. Thomas C, Dos'Santos T, Jones PA, and Comfort P. Reliability of the 30-15 intermittent fitness test in semiprofessional soccer players. *Int J Sports Physiol Perf* 11: 172-175, 2016.
78. Thomas C, Jones PA, and Comfort P. Reliability of the dynamic strength index in college athletes. *Int J Sports Physiol Perf* 10: 542-545, 2015.

79. Thorpe RT, Atkinson G, Drust B, and Gregson W. Monitoring Fatigue Status in Elite Team Sport Athletes: Implications for Practice. *Int J Sports Physiol Perf* 12, 2017.
80. Till K, Cobley S, O'Hara J, Cooke C, and Chapman C. Considering maturation status and relative age in the longitudinal evaluation of junior rugby league players. *Scand J Med Science Sports* 24: 569-576, 2014.
81. Till K, Cobley S, O'Hara J, et al. Retrospective analysis of anthropometric and fitness characteristics associated with long-term career progression in Rugby League. *J Sci Med Sport* 18: 310-314, 2015.
82. Till K, Scantlebury S, and Jones B. Anthropometric and physical qualities of elite male youth rugby league players. *Sports Med* 47: 2171-2186, 2017.
83. Treisman A. Preattentive processing in vision. *Comp Vis Grap Image Process* 31: 156-177, 1985.
84. Turner AN, Jones B, Stewart P, et al. Total Score of Athleticism: Holistic Athlete Profiling to Enhance Decision-Making. *Strength & Conditioning Journal* 41: 91-101, 2019.
85. Weakley J, Broatch J, O'Riordan S, et al. Putting the Squeeze on Compression Garments: Current Evidence and Recommendations for Future Research: A Systematic Scoping Review. *Sports Med* 52: 1141–1160, 2021.
86. Weakley J, Chalkley D, Johnston R, et al. Criterion Validity, and Interunit and Between-Day Reliability of the FLEX for Measuring Barbell Velocity During Commonly Used Resistance Training Exercises. *J Strength Cond Res* 34: 1519-1524, 2020.
87. Weakley J, Fernández-Valdés B, Thomas L, Ramirez-Lopez C, and Jones B. Criterion Validity of Force and Power Outputs for a Commonly Used Flywheel Resistance Training Device and Bluetooth App. *J Strength Cond Res* 33: 1180-1184, 2019.
88. Weakley J, Mann B, Banyard H, et al. Velocity-Based Training: From Theory to Application. *Strength Cond J* 43: 31-49, 2021.
89. Weakley J, McCosker C, Chalkley D, et al. Comparison of Sprint Timing Methods on Performance, and Displacement and Velocity at Timing Initiation. *J Strength Cond Res*, 2022.
90. Weakley J, Morrison M, García-Ramos A, et al. The validity and reliability of commercially available resistance training monitoring devices – A systematic review. *Sports Med* 51: 443-502, 2021.
91. Weakley J, Munteanu G, Cowley N, et al. The Criterion Validity and Between-Day Reliability of the Perch for Measuring Barbell Velocity During Commonly Used Resistance Training Exercises. *J Strength Cond Res,* 2022.
92. Weakley J, Till K, Darrall-Jones J, et al. Strength and conditioning practices in adolescent rugby players: relationship with changes in physical qualities. *J Strength Cond Res* 33: 2361-2369, 2019.
93. Weakley J, Till K, Sampson J, et al. The Effects of Augmented Feedback on Sprint, Jump, and Strength Adaptations in Rugby Union Players Following a Four Week Training Programme. *Int J Sports Physiol Perf*: 1205–1211, 2019.
94. Weakley J, Wilson K, Till K, et al. Show me, Tell me, Encourage me: The Effect of Different Forms of Feedback on Resistance Training Performance. *J Strength Cond Res* 34: 3157-3163, 2020.
95. Weakley J, Wilson K, Till K, et al. Visual feedback maintains mean concentric barbell velocity, and improves motivation, competitiveness, and perceived workload in male adolescent athletes. *J Strength Cond Res,* 2017.

96. Weakley JJ, Till K, Darrall-Jones J, et al. The influence of resistance training experience on the between-day reliability of commonly used strength measures in male youth athletes. *J Strength Cond Res* 31: 2005-2010, 2017.
97. Weakley JJS, Read DB, Fullagar HHK, et al. "How Am I Going, Coach?"-The Effect of Augmented Feedback During Small-Sided Games on Locomotor, Physiological, and Perceptual Responses. *Int J Sports Physiol Perform* 15: 677-684, 2020.
98. Weakley JJS, Till K, Read DB, et al. Jump Training in Rugby Union Players: Barbell or Hexagonal Bar? *J Strength Cond Res* 35, 2021.
99. Whitehead S, Weakley J, Cormack S, et al. The Applied Sports Science and Medicine of Netball: A Systematic Scoping Review. *Sports Med* 51: 1715-1731, 2021.
100. Wilson KM, de Joux NR, Head JR, et al. Presenting objective visual performance feedback over multiple sets of resistance exercise improves motivation, competitiveness, and performance. *Proc Hum Factors Ergon Soc* 62: 1306-1310, 2018.
101. Wilson KM, Helton WS, de Joux NR, Head JR, and Weakley JJ. Real-time quantitative performance feedback during strength exercise improves motivation, competitiveness, mood, and performance. *Proc Hum Factors Ergon Soc* 61: 1546-1550, 2017.
102. Woolford S, Polglaze T, Rowsell G, and Spencer M. Field Testing Principles and Protocols. in: *Physiological tests for elite athletes*. R Tanner, C Gore, eds. Champaign, IL, USA: Human Kinetics, 2012.

Figure 1. When deciding on a test, it is important to consider whether you can rank, monitor, and prescribe training for athletes with the collected data. While two of these outcomes may suffice, ideally, a test would have all three. An example of a commonly used test with all three considerations is the one repetition maximum (1RM) back squat. Coaches can prescribe with this data (particularly if this is combined with a load-velocity profile), use this information to help rank athletes as strength is an important physical quality across most sports, and monitor changes in strength over time as it has acceptable levels of reliability.

Figure 2. Types of validity and how they interact with each other.

Figure 3. A visual representation of validity and reliability.

Figure 4. A visual example of how “test error” can influence the interpretation of a performance score. Bars are the coefficient of variation (CV, the standard error of measurement as a percentage, shown hypothetically as 2%, 5%, and 10%).

Figure 5. Annotated example of change in 10m sprint performance in a single athlete (youth soccer player) across 12 weeks. The top figure illustrates the raw times (s) presented with the standard error of measurement (SEM), which is approximately 1.6% (24). The bottom figure demonstrates the corresponding test change score relative to the first testing occasion, presented with the adjusted SEM. The grey shaded region depicts the smallest worthwhile change (SWC) for 10-m sprint time in soccer players, which is said to be around 2% (24). A difference of 2% in 10-m time would allow a player to be ahead of an opponent over this distance in a one-on-one dual to win the ball (25).

Figure 6. An annotated example of change in 10-m sprint time in a group of athletes (youth soccer players). The variation around the point estimates (error bars with caps) represents the adjusted standard error of measurement (SEM), and the grey shaded region is the smallest worthwhile change (SWC) of 2% (See Figure 5 caption for further details). Also shown in this figure are compatibility limits of either 80% (thick grey line) or 90% (thin grey line). Depending on the certainty required for the measure, either option may be appropriate. This demonstrates how certain statistical choices can influence the interpretation of a change in test performance and the importance of showing uncertainty and practical importance. CI: Confidence Interval.

Figure 7. Recommendation of the order of tests when completing a testing battery in a single day (45). It should be noted that this order may differ if certain physical qualities are not assessed or if testing takes place across multiple days.

Figure 8. Changes in a back squat load-velocity profile of an athlete from baseline (black dots; grey line) and a load-velocity calculated during the warmup (red dots; light red line) three weeks later (93). Training had not been changed to record this information (i.e., 'invisibly monitored'). This information was then used to infer improvements in strength characteristics within a mesocycle. LVP: Load-velocity profile; V1RM: Velocity at one repetition maximum; Est. 1RM: estimated one repetition maximum). Light blue arrows demonstrate a change in the linear relationship.

Figure 9. Demonstrates the progressive reduction in cognitive load and decrease in processing time, establishing that there are 16 diamonds and 16 squares when color and grouping are strategically implemented.

Figure 10. The same data is presented in six different ways (sub-figure A-F), emphasizing and providing greater information through progressive layering of visual channels.

Figure 11. Z-score radar plot demonstrating the strengths and weaknesses of ‘Athlete 1’ in comparison to the ‘Academy Benchmark’. In this example, it can be observed that the athlete’s upper body strength is well above the benchmark, but further work in acceleration and maximal speed is required (46).

Figure 12. Figures 12A and B present athlete testing data through Z-scores and T-scores. In Figure 12B (T-scores), the yellow circle and number represents the athlete’s score out of 100, while the green and red values represent the highest and lowest scores from the cohort (84).

Figure 13. An image that emphasizes the inverse relationship between the time taken to present testing information and its usefulness to coaches.

Figure 14. A flowchart to help practitioners decide whether to use a certain test when assessing athletes.

Table 1. Overview and explanations of the different types of validity.

Table 2. Recommendations and considerations for improving test reliability in sport science.

Table 3. Between-day approximate coefficient of variations (%) for commonly used measures of physical capacity.

Table 4. Examples of additional outcomes that can be obtained from the addition of technology or a combination of other testing data in commonly used tests.

Table 5. Recommendations for presenting testing data to coaches and athletes.