Between-day reliability and usefulness of a fitness testing battery in youth sport athletes: Reference data for practitioners

Running Head: Reliability of a fitness testing battery in youth sport athletes

Thomas Sawczuk[1,2], Ben Jones[1,2,3,4,5], Sean Scantlebury[1,2], Jonathon Weakley[1,2,3], Dale Read[1,3], Nessan Costello[1,2,4], Joshua David Darrall-Jones[1,3], Keith Stokes[6] and Kevin Till[1,3,4]

[1] Institute for Sport, Physical Activity and Leisure, Leeds Beckett University, Leeds, United Kingdom

[2] Queen Ethelburga's Collegiate, Thorpe Underwood, York, United Kingdom

[3] Yorkshire Carnegie Rugby Club, Headingley Carnegie Stadium, Leeds, United Kingdom

[4] Leeds Rhinos Rugby Club, Headingley Carnegie Stadium, Leeds, United Kingdom

[5] The Rugby Football League, Red Hall, Leeds, United Kingdom

[6] University of Bath, Claverton Down, Bath, United Kingdom

Corresponding Author:

Thomas Sawczuk

Room G03, Macaulay Hall

Institute for Sport, Physical Activity and Leisure

Centre for Sports Performance

Leeds Beckett University, Headingley Campus

West Yorkshire

LS6 3QS

Phone: (0044) 7530945555

Email: t.sawczuk@leedsbeckett.ac.uk

Abstract

This study aimed to evaluate the between-day reliability and usefulness of a fitness testing battery in a group of youth sport athletes. Fifty-nine youth sport athletes (age = 17.3 ± 0.7 years) undertook a fitness testing battery including the isometric mid-thigh pull (IMTP), countermovement jump (CMJ), 5-40 m sprint splits, and the 5-0-5 change of direction test on two occasions separated by 7 days. Usefulness was assessed by comparing the reliability (typical error; TE) to the smallest worthwhile change (SWC). The TE was 5.5% for IMTP and 3.8% for CMJ. The TE values were 2.7%, 2.5%, 2.2%, 2.2% and 1.8% for the 5, 10, 20, 30 and 40 m sprint splits, and 4.1% (left) and 5.4% (right) for the 5-0-5 tests. SWC ranged from 1.1% to 6.1%. All tests were identified as having "good" or "acceptable" reliability. The IMTP and CMJ had "good" usefulness, all other tests had "marginal" usefulness.

Key words: Reliability, usefulness, fitness testing, strength, power, speed

Introduction

The importance of developing strength, power, speed and change of direction qualities to improve athletic performance and reduce injury risk in adolescent athletes has been highlighted regularly in recent years (Lloyd et al., 2016; Reilly, Williams, Nevill, & Franks, 2000; Young, 2006). This, in addition to the recent National Strength and Conditioning Association (NSCA) position statements indicating that resistance training is beneficial for the youth athlete, has resulted in an increase in the number of adolescents participating in structured strength and conditioning programmes (Faigenbaum et al., 2009; Lloyd et al., 2016). A number of these programmes take place at schools or colleges where adolescents are given scholarships based on their sporting prowess. A recent report by Ofsted has suggested that as many as 15% of current international athletes across a range of sports received a sports scholarship at some point during their school life, highlighting their importance (Ofsted, 2014). As part of the scholarships, it is common for coaches to use fitness testing batteries to regularly measure and monitor the physical characteristics of their athletes (Pyne, Spencer, & Mujika, 2014). However, little information is available regarding the between-day reliability of these tests, particularly in a school sport environment where athletes of different indoor and outdoor sports regularly train and test their physical capabilities together.

The between-day reliability of a test refers to its ability to produce consistent results from day to day (Hopkins, 2000). In order for coaches to be confident that changes in performance from a specific test are "real" and not due to the daily variation in the test, it is important that the test has good between-day reliability. Although good reliability of a test is necessary, in order for its results to be interpretable it is also important that it is sensitive enough to detect the smallest worthwhile change (SWC) in performance. This has been termed its "usefulness" (Hopkins, 2004) and is assessed by comparing a test's between-day reliability, or typical error (TE), to the SWC. To do this, the TE is usually converted into a factor of the SWC, which can be termed the "TE:SWC ratio". If the SWC is greater than a test's between-day reliability (i.e. TE:SWC ratio < 1) it is considered to have good usefulness. Conversely, if the SWC is smaller than its between-day reliability, (i.e. TE:SWC ratio > 1), its usefulness is said to be "marginal" (Hopkins, 2004). This information can be used to assess the length of time which may be required between tests in order for a clear change in performance to be seen. A number of tests of strength, power, speed and change of direction ability have had their between-day reliability and usefulness considered in recent times (Cormack, Newton, McGulgan, & Doyle, 2008; Darrall-Jones, Jones, Roe, & Till, 2015; De Witt et al., 2016; Gabbett, Kelly, & Sheppard, 2008; Haff, Ruben, Lider, Twine, & Cormie, 2015; Roe et al., 2016; Stewart, Turner, & Miller, 2014), however the reliability has been shown to vary between sports and cohorts, so a study considering the between-day reliability and usefulness of these tests in a group of school based adolescent athletes across multiple sports is warranted.

A recently conceptualised and novel single measure of strength is the isometric mid-thigh pull (IMTP). The IMTP is designed to mimic the second pull phase of the snatch and clean (Haff et al., 2005), and has shown to be strongly correlated with weightlifting performance (Beckham et al., 2013). It requires little technical expertise indicating it is suitable for athletes of all training ages with little familiarisation (Beckham, 2015). To date, the majority of IMTP reliability studies have shown force plates to be reliable measures, with intraclass correlation coefficients ranging from .92-.99 (De Witt et al., 2016; Haff et al., 2015), however given the high cost of the equipment, force plates are likely only available within universities or professional sports teams and not within a school environment. A lower cost alternative is to use a back dynamometer, the reliability of which (test-retest correlation, r = .91) has only been shown in older population groups (Coldwells, Atkinson, & Reilly, 1994). However, this study did not provide the typical error (TE) as a coefficient of variation (CV), limiting its practical use (Hopkins, 2000), so a study considering the between-day reliability and usefulness of the IMTP using a back dynamometer in youth sport athletes is warranted.

The countermovement jump (CMJ) has received considerable attention in recent years as a measure of neuromuscular power. Although the majority of this attention has come with regards to its use as a daily monitoring tool for neuromuscular power (McLean, Coutts, Kelly, McGuigan, & Cormack, 2010; Roe et al., 2016), CMJ height can still be used as a surrogate measure of lower body power similar to the vertical jump used in the NFL Combine (McGee & Burkett, 2003). As with the IMTP, the reliability of the CMJ has been shown extensively when using a force plate, with TE's ranging from 5.2-6.8% for jump height and 2.9-3.6% for flight time (Cormack et al., 2008; McLean et al., 2010; Roe et al., 2016). However, the less expensive Optojump system for measuring CMJ height has only once had its reliability confirmed (jump height TE = 2.2%) and this took place in a group of older, less well trained individuals (Glatthorn et al., 2011). It is therefore important to assess the between-day reliability and usefulness of the Optojump system as a measure of CMJ height in youth sport athletes.

Timing gates are frequently used to measure the linear sprint ability of athletes (Darrall-Jones et al., 2015; Duthie, Pyne, Ross, Livingstone, & Hooper, 2006; Young, McLean, & Ardagna, 1995). Within a cohort of rugby players the reliability of timing gates as a measure of linear sprint ability (TE = 1.3-3.1%) has been proven (Darrall-Jones et al., 2015), however the movement demands of rugby are different to those of netball (Chandler, Pinder, Curran, & Gabbett, 2014; Read et al., 2017), for example, thus differences in reliability may exist between sports. Furthermore, Darrall-Jones and colleagues (2015) found timing gates to have "marginal" usefulness, limiting the ability of the test to detect small changes in performance. In addition to possessing linear speed, it is important that athletes are able to move in a multi-directional manner (Sheppard & Young, 2006). As such, a test to determine the change of direction ability of youth athletes, for example the 5-0-5 test (Draper & Lancaster, 1985), should exist within a fitness testing battery. Previous studies have shown the 5-0-5 test measured using timing gates to be reliable in adult rugby league players (TE = 1.9%; Gabbett et al., 2008) and school PE students measured indoors (TE = 2.8%; Stewart et al., 2012). However, the 5-0-5 test is regularly assessed outdoors (Darrall-Jones, Jones, & Till, 2015; Gabbett et al., 2008) and differences in the between-day reliability of a test may be present in the same cohort in different conditions (i.e. indoor vs outdoor testing). Given the common use of timing gates to measure sprinting ability, a study involving numerous different sports, including both males and females, is required to establish the reliability and usefulness of timing gates for measuring speed in a multi-sport setting. Furthermore, establishing the reliability of the 5-0-5 test on an outdoor surface will provide useful information for coaches working with multi-sport youth athletes and will allow the usefulness of the test to be determined.

Despite a large number of sports scholarships across the globe, where general fitness testing protocols may be put in place, the majority of research considering the between day reliability of fitness testing protocols has taken place in individual sports (Cormack et al., 2008; Darrall-Jones et al., 2015; Gabbett et al., 2008). Consequently, the aim of this study was to assess the between-day reliability and usefulness of a fitness testing battery incorporating measures of strength, power, speed and change of direction ability in scholarship youth sport athletes.

Methods

Participants

Fifty-nine youth sport athletes (39 males, 20 females, age 17.3 ± 0.7 years, height 175.0 ± 17.4 cm, body mass 75.5 ± 14.0 kg) were recruited for this study from a local independent school in the United Kingdom. The athletes were part of the school's sport scholarship programme and had all previously competed at professional academy level or above, but were now club/school (n = 34), professional academy (n = 7), county/regional (n = 14) or international (n = 4) standard in their respective sports. The sports represented were basketball (n = 3), cricket (n = 5), football (n = 10), hockey (n = 9), netball (n = 10) and rugby (n = 22). Ethics approval was granted by the University Ethics Committee and written informed consent was provided by all participants and their parents prior to the study.

Research Design

In order to assess the between day reliability of this fitness testing battery, participants completed the tests on two separate occasions over a two-week period. Participants refrained from strenuous exercise in the 24 hours prior to each testing day, and training volume was standardised for the duration of the study, so that participants completed the same number of sessions in both weeks, in line with previous studies (Duthie et al., 2006; Stewart et al., 2012). These sessions consisted of strength and conditioning and technical training sessions, both of which were controlled for intensity between weeks. On days one and seven, subjects performed measures of strength via the IMTP and power via the CMJ. On days four and ten, field based measures of 40 m sprints to measure speed and the 5-0-5 test to measure change of direction ability were performed. On all testing days, the test inducing the greatest strain on the neuromuscular system was performed first in order to enhance the reliability of all maximal testing procedures (Harman, 2008). On days one and seven, this meant the CMJ was performed first; on days four and ten, the 5-0-5 change of direction test was performed first. Participants completed field based measures on either a 4G outdoor artificial grass playing surface (cricket, football and rugby) or an outdoor running track (basketball, hockey and netball), dependent on their sport. Ambient conditions were measured using a weather station (Davis Instruments Corporation, Hayward, USA). These are shown in Table 1. A standardised dynamic warm up including leg swings, lunges, squats was performed prior to each testing session. Participants had been familiarised to all tests earlier in the academic year.

INSERT TABLE 1 HERE

Protocols

The IMTP was performed using a modified back dynamometer (Takei Scientific Instruments Co., Niigata City, Japan). The modification increased the size of the base so that participants could stand with their feet shoulder width apart rather than hip width as necessitated by the original model. Participants were instructed to stand with their feet shoulder width apart and knees bent at 120-135˚ in line with previous studies (Beckham et al., 2013; Darrall-Jones et al., 2015). The bar was adjusted so that when they held it taut with a straight back and arms, it reached the middle of their thigh. Participants were instructed to pull directly upwards, keeping their feet flat on the platform and without leaning back. Two warm up pulls were performed at 50% and 75% of maximum, before three all out efforts were executed, each separated by a 3-minute rest.

The CMJ was performed using the Optojump system (Microgate, Bolzano, Italy). Jump height was reported in centimetres. Participants began with their legs fully extended and their feet at a self-selected width, with their hands on their hips. They were then instructed to squat down and jump as high as they could in a fluid, countermovement motion. The depth of the countermovement was self-selected. Participants were instructed to keep their legs extended in flight and to land with their legs straight. Two warm up jumps were completed, before three maximal efforts were executed with a 3-minute rest provided in between each repetition.

Linear sprint speed was assessed via 40 m sprint with split times taken at 5, 10, 20, 30 and 40 m using single beam timing gates (Brower Timing Systems, IR Emit, USA). The height of the timing gates was standardised at 1m in line with previous guidelines (Cronin & Templeton, 2008). The sprinting direction was standardised as north-west for all sessions. Participants were instructed to start 0.5 m behind the first timing gate and to start their sprints at a self selected time. Three maximal efforts were performed, each separated by 3-minutes rest. As part of their warm up, participants completed one practice sprint.

The 5-0-5 change of direction test was also measured using single beam timing gates (Brower Timing Systems, IR Emit, USA) and was performed after the 40 m sprints, following 5 minutes of rest. Participants began the test at a self selected time, sprinting 10 m in a south-easterly direction (i.e. opposite to the sprints) before planting their foot beyond a white line, turning 180˚ and sprinting back 5 m. The timing gates measured the time from 5 m before the line until they sprinted back through that point. Three maximal repetitions were completed with 3-minutes rest in between each effort. One practice test was completed as part of the warm up procedure.

Data Analysis

For all tests, the best of three efforts was taken for the between day reliability analysis, in line with previous studies (Darrall-Jones et al., 2015; Gabbett et al., 2008). Data were log transformed to allow the TE to be calculated as a CV (%), along with the SWC using a premade Microsoft Excel spreadsheet (Hopkins, 2015). The TE is calculated as follows:

$$TE = S_{diff}/\sqrt{2}$$

where S_{diff} is the standard deviation of the difference score between two trials (Hopkins, 2000). Back transformation of the log transformed data provided the TE as a percentage relative to the mean. Similar to previous studies, a CV of 5% or less was used to categorise a reliable variable (Darrall-Jones et al., 2015; Roe et al., 2016). The SWC was calculated as 0.2 x between-subject standard deviation, in line with the Cohen's *d* effect size principle, and expressed as a percentage of the mean in order to compare with the CV. The usefulness of the test was classified according to the Hopkins (2000) criteria: Good (CV < SWC; TE:SWC ratio < 1), OK (CV = SWC; TE:SWC ratio = 1) or Marginal (CV > SWC; TE:SWC ratio > 1).

Results

Table 2 shows the TE as a CV, SWC, TE:SWC ratio and usefulness of the tests. All sprint splits had less than 5% CV and so were considered reliable tests. The CMJ also showed good between day reliability. The IMTP and the 5-0-5 tests showed acceptable between day reliability. All timing gate measured tests' usefulness was considered "marginal", whereas the IMTP and CMJ tests were considered to have "good" usefulness.

INSERT TABLE 2 HERE

Discussion

The aim of this study was to assess the between-day reliability and usefulness of a fitness testing battery in a group of youth sport athletes of varying standards. The main finding is that all fitness testing protocols were shown to have good (< 5% TE) or acceptable (~5% TE) between-day reliability. A further finding of the study is that the strength and power tests showed "good" usefulness, whereas speed and change of direction tests using the timing gates showed "marginal" usefulness. As a consequence, the IMTP and CMJ tests are able to detect smaller changes in performance with greater certainty than the timing gate protocols.

A number of studies have considered the reliability of the IMTP using a force plate (De Witt et al., 2016; Haff et al., 2015), however to these authors' knowledge the only study to have considered a back dynamometer did not provide the TE of the test, and therefore did not provide practically useful results (Coldwells et al., 1994). Our results confirm the between-day reliability of the use of a back dynamometer to measure the IMTP strength of a youth sport athlete. The TE of 5.4% is slightly greater than, but still comparable with, the 1.7 and 3.1% values previously reported using force plates (Haff et al., 2015; James, Roberts, Haff, Kelly, & Beckman, in press). The difference in these figures is likely due to previously reported values using more expensive devices with greater sampling frequencies (e.g. force plates) and the different cohorts used. Furthermore, although it has previously been indicated that little familiarisation is required to perform the test (Beckham, 2015), the ability to consistently produce strength and power from this position is likely to require familiarisation to and training in Olympic weightlifting techniques. The 1.4% difference in the previously reported values (Haff et al., 2015; James et al., in press) provides support for this theory as the study by Haff and colleagues, which used participants who regularly performed "resistance training, including [Olympic] weightlifting movements" (Haff et al., 2015), showed greater reliability than the "recreationally active with ≥ 6 months resistance training exercise" participants used in James and colleagues' study (James et al., in press).

Our results demonstrate the reliability of the Optojump system for measuring power via CMJ height in adolescent youth sport athletes. The 2.8% TE here is similar to the 2.2% TE shown in an older, less trained age group by Glatthorn and colleagues (2011). Unlike the back dynamometer, this is much lower than those values reported in force plate studies (Cormack et al., 2008; Roe et al., 2016). It is suggested that this is because force plates calculate jump height differently to the Optojump system. Force plates calculate jump height using the velocity of the centre of mass at take off, whereas the Optojump system calculates flight time using the breaking of an infrared beam. No equation was provided by either previous study for their calculation of jump height from a force plate (Cormack et al., 2008; Roe et al., 2016), however both provided between-day reliability values for the flight time of the jump and our 2.8% between-day reliability value is comparable with the 3.3% (Cormack et al., 2008) and 2.6% (Roe et al., 2016) measurements previously quoted.

The results of the sprints show all splits to have good between-day reliability. Previous studies have either shown all sprint splits to be reliable (Darrall-Jones et al., 2015) or indicated that the shorter distances of 5 or 10 m are more unreliable than longer splits

(Earp & Newton, 2012). Our results are similar to those of Darrall-Jones and colleagues (2015) in that all splits were reliable, but also follow their trend of better reliability with increased sprint split distance. The difference in reliability between shorter (e.g. 5 m) and longer (e.g. 30 m) splits, however, was not sufficient for the shorter distances to provide a value greater than the 5% threshold set for good between-day reliability.

The results of the 5-0-5 change of direction test showed acceptable levels of reliability. Our values of 4.1% and 5.4% for left and right foot respectively are greater than the previously reported values of 2.8% in adult rugby league players (Stewart et al., 2012) and 1.9% in school physical education students, measured indoors (Gabbett et al., 2008). This variation in the studies' results may be due to differences in the techniques used to measure speed (e.g. single beam electronic timing gates vs the dual beam electronic timing gates used by Gabbett and colleagues (2008)) and the differences between testing outdoors or on an indoor running surface, as used by Stewart and colleagues (2012). Although there is little difference in the ambient conditions shown in Table 1, the slight reduction in reliability is likely due to the inherent variability in weather and ground conditions for outdoor field testing, and thus should be a consideration for those involved in testing young athletes. This is also shown by the difference in reliability between the two previous studies (Gabbett et al., 2008; Stewart et al., 2012). The between-day difference in sprint time was lowest in the study of Stewart et al. (2014), where testing was undertaken indoors, despite the more precise timing gates used by Gabbett and colleagues (2008), where testing was undertaken outdoors.

With regards to the usefulness of the tests, the IMTP and CMJ tests were identified as useful tests (TE:SWC ratio < 1). However, despite showing good reliability, the sprint splits and 5-0-5 change of direction test were identified as having "marginal" usefulness (TE:SWC ratio > 1). In order for a practitioner to assess whether a "real" change in performance has occurred, it is important to plot the change in performance ± TE against the SWC. If the TE remains outside the SWC, it can be postulated that a change has occurred with 75% probability, however if the TE crosses the SWC, the changes are deemed unclear (Hopkins, 2000). Figure 1 shows a practical example of this, plotting a sample change in performance for the 10 and 20 m sprint splits of a player, using the SWC and TE values reported in this paper. It is due to this method of interpreting results that the usefulness of a test is important. In order for the change in performance to be considered "real" with 75% confidence, it must be greater than or equal to the SWC + TE. In the case of the IMTP and CMJ tests, this means that a change of 1.9 and 1.7 times the SWC would be required respectively; however for the 5-0-5 test this can rise to as much as four times the SWC. As a consequence, a much greater improvement in performance is needed for practitioners to be confident that a "real" change has occurred. Given an average 5-0-5 performance of 2.5 s and an SWC of 1.8% (0.05 s), a change in performance of 0.20 s would be required to be sure with 75% probability that a "real" change has occurred. This may be an unrealistic expectation if the test is to be used on a regular (e.g. monthly/quarterly) basis, rather than a longer-term basis (e.g. bi-annually/annually). In this situation, practitioners are advised to use the test less regularly (e.g. bi-annually/annually) so that performance changes are greater and therefore more certain, or accept that there will be a large element of uncertainty in results if testing occurs regularly and performance changes are smaller.

INSERT FIGURE 1 HERE

Although this study has shown the reliability of a fitness testing battery in male and female youth sport athletes with uneven representations from a number of sports, it is important to understand the limitations inherent in the testing protocols. It could be argued that not splitting the tests by sex is a limitation of the study, however in the only study to date to compare the reliability of a fitness test between sex, no significant differences in reliability were found (Augustsson et al., 2009). The multi-sport nature of the study could be seen as a limitation. It has previously been noted that the movement demands of the sports are different (Chandler et al., 2014; Read et al., 2017), however this study shows that within a multi-sport environment, this fitness testing battery remains reliable. Future research may wish to assess whether there are differences in reliability between sports, however given previous research that shows no differences in reliability between sex (Augustsson et al., 2009), it is possible that the differences between sports will be negligible. The use of two different surfaces for the sprint based tests (outdoor athletics track and 4G artificial grass) could also be seen as a limitation, however this could happen in practice if teams were to perform sprint testing on their normal playing surface. Furthermore, as the surface remained constant between days, it should not have had an impact on the results seen.

Finally, in order for this testing battery to be considered complete, a marker of aerobic fitness should also be included. Along with its performance related benefits, aerobic fitness has been associated with reduced injury risk in adolescents (Brenner, 2007; Difiori et al., 2014). It is a limitation of this study that a test of aerobic fitness was not included, however both the Yo-Yo Intermittent Recovery Test (Bangsbo, 1994; Bangsbo, Iaia, & Krustrup, 2008) and the 30-15 Intermittent Fitness Test (Buchheit & Rabbani, 2014) have had their reliability and sensitivity to training confirmed in similar populations and conditions to this study (Buchheit & Rabbani, 2014; Deprez et al., 2014). Consequently, the authors decided not to include either of these tests in this study.

In conclusion, this study has shown the reliability and usefulness of a fitness testing battery aimed at monitoring strength, power and speed qualities in youth sport athletes. The IMTP and CMJ were shown to be both reliable and have good usefulness. The sprint splits and 5-0-5 test were shown to be reliable but had marginal usefulness. To this end, the IMTP and CMJ are able to detect the smaller changes in performance with greater certainty than the sprint splits and 5-0-5 test. It is recommended that either the Yo-Yo Intermittent Recovery Test Level 1 or 30-15 Intermittent Fitness Test is added to ensure the battery provides a complete understanding of the athlete's physical capabilities.

References

Augustsson, S. R., Bersas, E., Thomas, E. M., Sahlberg, M., Augustsson, J., & Svantesson, U. (2009). Gender differences and reliability of selected physical performance tests in young women and men. *Advances in Physiotherapy*, *11*, 64–70.

Bangsbo, J. (1994). *Fitness training in football: A scientific approach*. Copenhagen University: August Krogh Institute.

Bangsbo, J., Iaia, F. M., & Krustrup, P. (2008). The Yo-Yo Intermittent Recovery Test: A useful tool for evaluation of physical performance in intermittent sports. *Sports Medicine*, *38*, 37–51. http://doi.org/10.2165/00007256-200838010-00004

Beckham, G. K. (2015). *The effect of various body positions on performance of the isometric mid-thigh pull*.(PhD Thesis). East Tennessee State University. Retrieved from http://dc.etsu.edu/etd/2544

Beckham, G., Mizuguchi, S., Carter, C., Sato, K., Ramsey, M., Lamont, H. S., … Stone, M. H. (2013). Relationships of isometric mid-thigh pull variables to weightlifting performance. *Journal of Sports Medicine and Physical Fitness*, *53*, 573–581.

Brenner, J. S. (2007). Overuse injuries, overtraining, and burnout in child and adolescent athletes. *Pediatrics*, *119*, 1242–1245. http://doi.org/10.1542/peds.2007-0887

Buchheit, M., & Rabbani, A. (2014). The 30-15 Intermittent Fitness Test versus the Yo-Yo Intermittent Recovery Test Level 1: Relationship and sensitivity to training. *International Journal of Sports Physiology and Performance*, *9*, 522–524. http://doi.org/10.1111/j.1399-0012.2012.01641.x

Chandler, P. T., Pinder, S. J., Curran, J. D., & Gabbett, T. J. (2014). Physical demands of training and competition in collegiate netball players. *Journal of Strength and Conditioning Research*, *28*, 2732–2737. http://doi.org/10.1519/1533-4287(2005)19<202

Coldwells, A., Atkinson, G., & Reilly, T. (1994). Sources of variation in back and leg dynamometry. *Ergonomics*, *37*, 79–86.

Cormack, S. J., Newton, R. U., McGulgan, M. R., & Doyle, T. L. A. (2008). Reliability of measures obtained during single and repeated countermovement jumps. *International Journal of Sports Physiology and Performance*, *3*, 131–144.

Cronin, J. B., & Templeton, R. L. (2008). Timing light height affects sprint times. *Journal of Strength and Conditioning Research*, *22*, 318–20. http://doi.org/10.1519/JSC.0b013e31815fa3d3

Darrall-Jones, J. D., Jones, B., Roe, G., & Till, K. (2015). Reliability and usefulness of linear sprint testing in adolescent rugby union and league players. *Journal of Strength and Conditioning Research*, *30*, 1359–1364. http://doi.org/10.1519/JSC.0000000000001233

Darrall-Jones, J. D., Jones, B., & Till, K. (2015). Anthropometric and physical profiles of English academy rugby union players. *Journal of Strength and Conditioning Research*, *29*, 2086–2096.

De Witt, J. K., English, K. L., Crowell, J. B., Kalogera, K. L., Guilliams, M. E., Nieschwitz, B. E., … Ploutz-Snyder, L. L. (2016). Isometric mid-thigh pull reliability and relationship to deadlift 1RM. *Journal of Strength and Conditioning Research*, *epub ahead of print*. http://doi.org/10.1519/JSC.0000000000001605

Deprez, D., Coutts, A. J., Lenoir, M., Fransen, J., Pion, J., Philippaerts, R., & Vaeyens, R. (2014). Reliability and validity of the Yo-Yo intermittent recovery test level 1 in young soccer players. *Journal of Sports Sciences*, *32*, 903–910. http://doi.org/10.1080/02640414.2013.876088

Difiori, J. P., Benjamin, H. J., Brenner, J., Gregory, A., Jayanthi, N., Landry, G. L., & Luke, A. (2014). Overuse injuries and burnout in youth sports: A position statement from the American Medical Society for Sports Medicine. *Clinical Journal of Sport Medicine*, *24*, 3–20. http://doi.org/10.1097/JSM.0000000000000060

Draper, J. A., & Lancaster, M. G. (1985). The 505 test: A test for agility in the horizontal plane. *Australian Journal of Science and Medicine in Sport*, *17*, 15–18.

Duthie, G. M., Pyne, D. B., Ross, A. A., Livingstone, S. G., & Hooper, S. L. (2006). The reliability of ten-meter sprint time using different starting techniques. *Journal of Strength and Conditioning Research*, *20*, 246–251. http://doi.org/10.1519/R-17084.1

Earp, J. E., & Newton, R. U. (2012). Advances in electronic timing systems: Considerations for selecting an appropriate timing system. *Journal of Strength and Conditioning Research*, *26*, 1245–1248.

Faigenbaum, A. D., Kraemer, W. J., Blimkie, C. J. R., Jeffreys, I., Micheli, L. J., Nitka, M., & Rowland, T. W. (2009). Youth resistance training: Updated position statement paper from the National Strength and Conditioning Association. *Journal of Strength and Conditioning Research*, *23*, S60–S79. http://doi.org/10.1519/JSC.0b013e31819df407

Gabbett, T. J., Kelly, J. N., & Sheppard, J. M. (2008). Speed, change of direction speed, and reactive agility of rugby league players. *Journal of Strength and Conditioning Research*, *22*, 174–181. http://doi.org/10.1519/JSC.0b013e31815ef700

Glatthorn, J. F., Gouge, S., Nussbaumer, S., Stauffacher, S., Impellizzeri, F. M., & Maffiuletti, N. A. (2011). Validity and reliability of Optojump photoelectric cells for estimating vertical jump height. *Journal of Strength and Conditioning Research*, *25*, 556–560.

Haff, G. G., Carlock, J. M., Hartman, M. J., Kilgore, J. L., Kawamori, N., Jackson, J. R., … Stone, M. H. (2005). Force-time curve characteristics of dynamic and isometric muscle actions of elite women Olympic weightlifters. *Journal of Strength and Conditioning Research*, *19*, 741–748.

Haff, G. G., Ruben, R. P., Lider, J., Twine, C., & Cormie, P. (2015). A comparison of methods for determining the rate of force development during isometric midthigh clean pull. *Journal of Strength and Conditioning Research*, *29*, 386–395.

Harman, E. (2008). Principles of test selection and administration. In T. R. Baechle & R. W. Earle (Eds.), *Essentials of Strength and Conditioning* (3rd ed., pp. 238–246). Champaign, IL: Human Kinetics.

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, *30*, 1–15. http://doi.org/10.2165/00007256-200030050-00006

Hopkins, W. G. (2004). How to interpret changes in an athletic performance test. *Sportscience*, *8*, 1–7.

Hopkins, W. G. (2015). Spreadsheets for analysis of validity and reliability. *Sportscience*, *19*, 36–42.

James, L. P., Roberts, L. A., Haff, G. G., Kelly, V. G., & Beckman, E. M. (2017). The validity and reliability of a portable isometric mid-thigh clean pull. *Journal of Strength and Conditioning Research*, *31*, 1378–1386. http://doi.org/10.1519/JSC.0000000000001201

Lloyd, R. S., Cronin, J. B., Faigenbaum, A. D., Haff, G. G., Howard, R., Kraemer, W. J., … Oliver, J. L. (2016). National Strength and Conditioning Association Position Statement on Long Term Athletic Development. *Journal of Strength and Conditioning Research*, *30*, 1491–1509.

McGee, K. J., & Burkett, L. N. (2003). The National Football League Combine: A reliable predictor of draft status? *Journal of Strength and Conditioning Research*, *17*(1), 6–11. http://doi.org/10.1519/1533-4287(2003)017<0006:TNFLCA>2.0.CO;2

McLean, B. D., Coutts, A. J., Kelly, V., McGuigan, M. R., & Cormack, S. J. (2010). Neuromuscular, endocrine, and perceptual fatigue responses during different length between-match microcycles in professional rugby league players. *International Journal of Sports Physiology and Performance*, *5*, 367–383.

Ofsted. (2014). *Competitive school sport - Summary report*. Loughborough, United Kingdom. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/379986/National_20Governing_20Bodies_20of_20Sport_20survey_2C_20Competitive_20school_20sport.pdf

Pyne, D. B., Spencer, M., & Mujika, I. (2014). Improving the value of fitness testing for football. *International Journal of Sports Physiology and Performance*, *9*, 511–514.

Read, D., Jones, B., Phibbs, P., Roe, G., Darrall-Jones, J. D., Weakley, J., & Till, K. (2017). Physical demands of representative match play in adolescent rugby union. *Journal of Strength and Conditioning Research*, *31*, 1290–1296. http://doi.org/10.1519/JSC.0000000000001600

Reilly, T., Williams, A. M., Nevill, A., & Franks, A. (2000). A multidisciplinary approach to talent identification in soccer. *Journal of Sports Sciences*, *18*, 695–702. http://doi.org/10.1080/02640410050120078

Roe, G., Darrall-Jones, J. D., Till, K., Phibbs, P., Read, D., Weakley, J., & Jones, B. (2016). Between-day reliability and sensitivity of common fatigue measures in rugby players. *International Journal of Sports Physiology and Performance*, *11*, 581–586. http://doi.org/10.1123/ijspp.2015-0413

Sheppard, J. M., & Young, W. B. (2006). Agility literature review: Classifications, training and testing. *Journal of Sports Sciences*, *24*, 919–932. http://doi.org/10.1080/02640410500457109

Stewart, P. F., Turner, A. N., & Miller, S. C. (2012). Reliability, factorial validity and interrelationships of five commonly used change of direction speed tests. *Scandinavian Journal of Medicine & Science in Sports*, *24*, 500–506. http://doi.org/10.1111/sms.12019

Young, W. B. (2006). Transfer of strength and power training to sports performance. *International Journal of Sports Physiology and Performance*, *1*, 74–83. http://doi.org/10.1123/ijspp.1.2.74

Young, W., McLean, B., & Ardagna, J. (1995). Relationship between strength qualities and sprinting performance. *Journal of Sports Medicine and Physical Fitness*, *35*, 13–19.

Table 1: Ambient conditions for field testing for each sport.

Sport	**Temperature (˚C)**	**Humidity (%)**	**Air Pressure (hPa)**	**Wind (m/s)**	**Description**
Basketball Day 1	18	54	1008	North (4.6)	Scattered Clouds
Basketball Day 2	12	82	1010	South (3.6)	Haze
Cricket Day 1	17	58	1008	North (4.6)	Scattered Clouds
Cricket Day 2	12	82	1010	South (3.6)	Haze
Football Day 1	19	46	1013	ENE (4.6)	Scattered Clouds
Football Day 2	13	58	1020	West (1.5)	Partly Cloudy
Hockey Day 1	18	54	1008	North (4.6)	Scattered Clouds
Hockey Day 2	12	82	1010	South (3.6)	Haze

Netball Day 1	18	54	1008	North (4.6)	Scattered Clouds
Netball Day 2	14	69	1010	SSE (3.1)	Overcast
Rugby Day 1	17	53	1013	ENE (4.6)	Scattered Clouds
Rugby Day 2	12	67	1020	West (2.1)	Partly Cloudy

Note: Running direction was standardised as North West for sprints and South East (before the change of direction) for the 5-0-5 test.

Table 2: Summary of day 1 and day 2 raw values, TE as a CV (90% confidence intervals in brackets) and SWC as percentages, TE:SWC ratio and usefulness (Hopkins, 2000) for each test.

Test	**Day 1**	**Day 2**	**TE as a CV (%)**	**SWC (%)**	**TE:SWC ratio**	**Usefulness**
IMTP (kg)	170.6 ± 45.5	170.9 ± 46.4	5.5 (4.5 - 6.9)	6.1	0.9	Good
CMJ (cm)	34.4 ± 5.9	34.4 ± 6.4	2.8 (2.4 - 3.3)	3.9	0.7	Good
5 m (s)	1.08 ± 0.06	1.06 ± 0.06	2.7 (2.0 - 4.0)	1.1	2.5	Marginal
10 m (s)	1.82 ± 0.09	1.78 ± 0.11	2.5 (2.1 - 3.2)	1.1	2.3	Marginal
20 m (s)	3.19 ± 0.17	3.10 ± 0.19	2.2 (1.9 - 2.8)	1.1	2.0	Marginal
30 m (s)	4.45 ± 0.28	4.37 ± 0.28	2.2 (1.8 - 2.7)	1.3	1.7	Marginal
40 m (s)	5.75 ± 0.38	5.68 ± 0.42	1.8 (1.5 - 2.3)	1.4	1.3	Marginal
5-0-5 L (s)	2.54 ± 0.21	2.50 ± 0.22	4.1 (3.4 - 5.4)	1.7	2.4	Marginal
5-0-5 R (s)	2.49 ± 0.20	2.52 ± 0.25	5.4 (4.4 - 7.0)	1.8	3.0	Marginal

IMTP = Isometric Mid Thigh Pull; CMJ = Countermovement jump; L = Left; R = Right

Figure 1: An example of the change in performance of an athlete over two tests using our reliability data. Data are percentage change in performance (± TE as a CV as error bars). The shaded grey area represents the SWC, which is the same for both tests. Although the magnitude of improvement in performance is the same, the difference in the TE results in the error bar overlapping the SWC for the 10 m split, leading to an inability to describe the changes as "real" with 75% confidence. The error bar for the 20 m split does not overlap the SWC however, resulting in a clear improvement in performance, with 75% certainty (Hopkins, 2000).