The efficacy of core stability assessment as a determiner of performance in dynamic balance and agility tests

KIERAN NEIL MCCARTNEY¹, JACKY FORSYTH²

¹Open University and The North West Regional College, United Kingdom
²Staffordshire University, United Kingdom

ABSTRACT

Aim: The aim of this study was to investigate if tests used to assess core stability could be used to determine success in physiological tests applied to assess dynamic balance and agility for a young active population.

Methods: Pearson’s r correlation coefficient was used to assess the relationship between the core stability tests and the dynamic balance and agility tests. Evaluation of the tests was established using Cronbach’s coefficient of variance as part of intra-rater reliability tests. An analysis of 18 active college aged students was conducted (males: n= 13, females: n= 5). The mean ± SD age for males was 19.2 years ± 3.22 years and for females was 19.4 years ± 1.14 years. Conclusion: The results indicate that there is no significant relationship between tests that assess core stability and tests conducted to assess dynamic balance in active young adults. With the exception of the abdominal flexion test, no significant relationship exists between the remaining core stability tests and agility T-Test. Core stability is not a determinant of balance and agility. Key words: CORE STABILITY, BALANCE AND AGILITY.

Cite this article as:
INTRODUCTION

In recent years there has been a growing interest in the role that core stability has in improving sporting performance, fitness, and in preventing injury (Granacher, Lacroix, Roettger, Gollhofer and Muehlbauer, 2014; Kibler, Press & Sciascia, 2006; Singh, Sharma & Hussain, 2012). The purpose of the studies, which have been conducted to date, has been to investigate the relationship between core stability and functional movement, and/or to focus on how core stability training can be incorporated into an athlete's training regime. These studies have ecological validity, since the researchers investigated functional movements, which were designed to replicate day-to-day activities or specific sporting scenarios.

Providing a clear definition of core stability has resulted in indeterminate conclusions. Several researchers have attempted to determine not only the factors that influence core stability but also how core stability can be measured. Despite the variations that exist in the plethora of research in this area, there is agreement that a difference between core strength and core stability is evident (add some references here). Core strength involves training the central muscles to withstand considerable force demands and is useful in contact sports such as weight-training, judo and soccer, as highlighted by Zemková (2014). In contrast, core stability refers to the ability of the joints and specific muscle groups to facilitate movement within safe structural limits. Furthermore, core stability has a relationship with balance and agility. For example, Granacher et al (2014) demonstrated that the ability to prevent injury as a result of slips trips or falls in the elderly was reduced when the core stabilising muscles were engaged. This research suggests that core stability and balance are related to one another.

Research by Araujo, Cohen and Hayes' (2015) and Petersen and Nittinger's (2014) indicates that the stabilising muscles of the core muscles aid in dynamic movement in sporting and non-sporting environments. Therefore, core stability involves the effective utilization of muscular potential. This is a concept outlined in Lee's (2005) study, where the author highlighted the importance of the neuromuscular responses due to autonomic activation of the joint stabilizing muscles in order to address unexpected challenges associated with a loading response (i.e. balance) and the body's ability to respond to rapid and/or planned directional changes (i.e. agility). Consequently, the common theme that is eminent throughout much of the research is that core stability has an impact upon balance and agility (Araujo, Cohen and Hayes', 2015; Lee's, 2005 and Petersen and Nittinger's, 2014).

Several studies have sought to demonstrate the importance of core stability in relation to athletic performance, such as tennis, and netball (de Villiers and Venter, 2014; McCurdy, Smart, Pankey and Walker, 2014 and Sharrock, Cropper, Mostad, Johnson and Malone, 2011). However, one of the challenges evident in this research are the inconsistencies when determining what role the stabilising muscles of the abdominal region have in relation to balance and agility. This may, in part, be as a result of the use of core strength and core stability as interchangeable terms, where in fact there are clear differences as already outlined (Araujo, Cohen and Hayes, 2015 and Zemková, 2014).

One important area of enquiry that has not yet yielded results is the relationship between the tests applied for core stability and their relationship with associated balance and agility tests. In order to clarify this issue, the relationship between tests for core stability and tests for balance and agility could be explored. In order to illuminate the relationship between core stability and associated balance and agility tests, reliable and valid tests should be chosen.

The abdominal flexion test; The back extension test and the side balance tests (Left and Right) are all core
stability tests. Reliability coefficients of 95% CI: .99% for the side balance tests (left and right) and further 95% CI: .98 and .97 for the back extension test and abdominal flexion test, respectively, have been demonstrated by McGill, Childs and Liebenson (1999). More recently Waldhelm and Li (2012) reported intra-rater reliabilities of r=.66, .79, .74 and .96 and confidence intervals of 95% CI: .01–.89, .38–.93, .30–.92, .87–.99 for the back extension test, abdominal flexion test and side balance test, respectively.

Regarding tests to measure agility, Raya, Gailey, Gaunaurd, Jayne, Campbell, Gagne, Tucker's (2013) applied the intra-class correlation coefficient as part of an analysis of completion times for the agility T-test (T-Test), the Illinois agility test (IAG) and the Edgren Side Step Test (ESST) that revealed results of 95% CI: 98, .99 and .92 for each test respectively. Paule, Madole, Garhammer, Lacourse and Rozenek (2000) offer corroborating findings in relation to the T-Test as a measure of agility establishing intra-class correlation coefficient that ranged between 95% CI: .94 - .98, the range present owing to the test results evident for three trials. Given the comparison between Raya et al (2013) and Paule et al's (2000) findings in relation to the T-Test, this test has been determined as a more reliable test of agility than the ESST and the IAG.

The star excursion balance test has also been found to be reliable for assessing dynamic balance. Gribble, Kelly, Refshauge and Hiller (2013); Hyong and Kim, (2014) and Munro and Herrington (2010) assessed the inter and intra-rater reliability of the star excursion balance test using the intra-class correlation coefficient (ICC) and found the test to have moderate to high levels of reliability when assessing dynamic balance. Munro and Herrington's (2010) research revealed test-retest reliability for all reach directions, in the star excursion balance test, in relation to intra-class correlation coefficients ranging from 95% CI: .84 to .92. Similar ranges were revealed in Gribble et al's (2013) study that also considered the efficacy of normalising the measurements in the star excursion balance test by using the scores captured from test as a percentage of relevant leg length (eg length was measured from the anterior superior iliac spine (ASIS) to the medial malleolus). Reliability for normalised measures proved higher where the intra-class correlation coefficient revealed 95% CI: .89 to .94 compared with non-normalised ICC measurements of 95% CI: .86 to .92. Similar findings were also produced during Hyong and Kim's (2014) research that outlined inter-rater reliability. The ICC values for all directions ranged from 95% CI: .83 to .93. Gribble et al's (2013), Hyong and Kim (2014) and Munro and Herrington's (2010) research were limited to a study of the star excursion balance test itself and not against other balance tests. The wider array of evidence from intra-class correlation coefficient measurements from these researchers that ranges from 95% CI: .83 to .93 provides stronger evidence for this test’s use.

The rationale of the project is to assess whether three core stability tests, the abdominal flexion test, the back extension test and the side balance test (both left and right sides) outlined by Waldhelm and Li (2012), correlate with performance in the star excursion balance test and the agility T-Test for a young and active population.

MATERIALS AND METHODS

Participants
How recruited? Mean (± SD) age of 22 males was 19.54±2.72 years, and for females (n= 5) was 19.40±1.14 years. Participants spent a mean of 2.77 ± 0.80 days per week training in their sport-specific area and a mean of 1.85 ± 1.02 days per week training in non-sport specific areas. Participants were encouraged to avoid eating a large meal one hour prior to testing, and to avoid fluids 30 min prior to testing. Participation in sporting activities varied substantially, including, cricket, rugby, soccer, gaelic football, kickboxing, hockey, judo, running, and gym.
Participants were excluded from the study if they indicated any medical conditions that impeded their performance or exacerbated their condition. Inclusion criteria for the study involved being active in either a sport or other activity. The participants provided informed consent approved by the Institutional Review Board prior to the collection of data.

**Procedures**

A design was implemented to assess inter-rater reliability and test-retest reliability. All participants who attended the sessions had the measurements of their tests recorded independently by two assessors and three assessors where the star excursion balance test was assessed in accordance with Gribble et al's (2013) assertions. Upon completion of each test participants were allowed a 3-min rest period to account for depletion of adenosine triphosphate and phosphocreatine, as per the American College of Sports Medicine (2006) guidelines.

On each testing day a random 10% of participants were invited to retake the assessments for the purpose of assessing the coefficient of variance in a test re-test period. How did randomisation occur? There were three testing periods therefore a total of n=10 participants (three from each testing period) were invited to participate. The random 10% of participants who were invited back for re-testing repeated the tests one week after their initial testing on the same day of the week and time as well as in the same order, following the same procedures and protocols outlined in the initial testing period. A period of one week was chosen to account for the influence of the learning effect that was also considered within Munro and Herrington's (2010) research relating to the star excursion balance test and the T-Test in Raya et al's (2013) study.

The core stability assessments were conducted as a battery of tests together and non-sequentially. The sequencing of the participants' completion of the core stability tests, agility and dynamic balance test was conducted in a randomised order. Randomised testing was conducted in order to account for the influence that fatigue in one test would have upon another as a result of the differing energy pathways required to conduct each test (American College of Sports Medicine, 2006).

The best of three attempts for the core stability, agility and dynamic balance tests were recorded. The participant's leg length, age, activity levels, sporting activities as well as food and fluid consumption prior to the tests were recorded. A 10-min warm-up was performed where participants engaged in a standardised pulse raising activity that involved jogging around a set area followed by dynamic stretching and mobility exercises.

**Core stability tests**

The abdominal flexion test was performed on a Fitness Mat using a GemRed Digital Angle Rule to calculate the measurements of the required angles. This test was designed to assess the anterior musculature of the rectus abdominis. The participants were offered practice trials where the angle of their back and positioning of the hips and knees could be judged in relation to the test requirements. Participants were encouraged to maintain the position for as long as possible. Testers continuously inspected maintenance of the angles required, i.e. 60° reclined back position and both the hips and knees bent at 90°. A stop watch was used to record each tester's results, recorded to the nearest 0.01 s.

The back extension test was performed on a 750-g adjustable incline bench (Pulse Fitness, Antrim). This witnessed a departure from Biering-Sorensen's (1984) original protocol, in which it was suggested that participants be restrained against an exercise bench using belts to ensure the participants could conduct the exercise properly. Instead, the participant's bodyweight was stabilised by an assistant who, using their hands,
secured the ankle position of the participant. This change in procedure facilitated release from the exercise so that participants could dismount the apparatus more safely given its closer proximity to the floor. Similarly to the abdominal flexion test, outlined above, participants were encouraged to practise the movement of the assessment prior to commencing the test. Participants were encouraged to maintain the position for as long as possible. Testers regularly reviewed the participant's positioning and stopped the timer when the participant dipped below the edge of the bench or if the participant reached exhaustion. A stop watch was used to record each tester's results, recorded to the nearest 0.01 s.

A further core stability test involved an assessment of the lateral musculature (left and right side). Participants were encouraged to practise the movement of the assessment prior to commencing the test. Participants were asked to maintain the position for as long as possible but were not offered verbal or other forms of encouragement during the test. Timing for the test was terminated when the participants could no longer maintain either the left or right side balance. A stop watch was used to record each tester's results, recorded to the nearest 0.01 second.

**Dynamic Balance Test**

The original protocol for the star excursion balance test outlined by Brumitt (2010) has, as a result of further research by Coughlan, Fullam, Delahunt, Gissane, Caulfield (2012) and Shaffer, Teyhen, Lorenson, Warren, Koreerat, Straseske and Child (2013), been narrowed to encompass assessment of only the anterior, posteromedial and posterolateral positions. Coughlan et al (2012) and Shaffer et al (2013) indicate that these three positions produce equivalency in results when compared with assessing all six points of the original test. Given these revisions, the following procedure for the dynamic balance test were adopted: Participants were, prior to taking the test, required to have their leg length measured in accordance with methods outlined by Coughlan, Delahunt, O'Sullivan, Fullam, Green and Caulfield (2014) and Gribble and Hertel (2003). The methods outlined by the aforementioned researchers dictate that consideration be given to a method that normalises the results of the test given that leg length varies between individuals. In order to account for variation between leg lengths in the star excursion balance test, Gribble et al's (2013); Munro and Herrington's (2010); Sabharwal and Kumar's (2008) methodologies were adopted. Here participants' legs were measured from the anterior superior iliac spine to the medial malleolus. Participants were offered three attempts at the test; the best result of the three tests was recorded.

An Irwin tape measure was used to determine distance between the starting and ending point of leg reach during the test. Participants did not wear footwear during the performance of the test. Athletic tape was used to construct the star excursion balance test in accordance with Brumitt's (2010) original procedure. In a modification, to ensure greater accuracy, three testers monitored the placement of the participant's bare foot during the test, reflective of Gribble et al's (2013) enhanced protocol. In addition, the difference between the first, second and third attempts were noted using numbers alongside marks on the tape. The tape was replaced between participants' attempts.

**Agility test**

Completion of the T-Test followed the protocol established by Seminick (1990). Participants were afforded the opportunity of three attempts with the highest score on the test used for analysis. Completion times were monitored by two assistants, who used stop watches to record completion times to the nearest 0.01 s. No incentives or verbal encouragement were offered, during any of the test procedures.

The test period took place over three consecutive weeks in the same location, on the same day and at the same time. Temperature (ºC) and relative humidity (%) were recorded on each testing day. The temperature
was recorded using a liquid thermometer and humidity using a hygrometer.

**Analysis**

Upon completion of the test period the results were analysed using SPSS for Windows (version 20.00; SPSS Inc.). Descriptive statistics were collated (mean ± SD). All data were initially evaluated for linearity and normality, applying the Sapiro-Wilk test given the sample size was less than 50.

Pearson’s Correlation Coefficient was used to assess the relationship between the three core stability tests, the T-Test and the star excursion balance test. Pearson’s correlation coefficient was complemented with the use of the ICC given Pearson’s correlation coefficient strength lies in the determination of the strength of a linear relationship and not agreement.

The ICC was applied in order to assess both inter-rater reliability and test-retest reliability. Inter-rater reliability refers, in this research, to the variation of measurements gathered by two or more assessors within the same group of participants (O’Donoghue, 2012). The inter-rater reliability was calculated using the ICC with 95% confidence intervals for each of the variables being assessed. The ICC was also used to assess the reliability of the test re-test results that were conducted one week after the test period had been completed. The primary difference between inter-rater reliability and re-test reliability is that the participants were tested more than once.

As part of this analysis Kline’s (1999) scale of reliability will be used where $\alpha \geq 0.9$ is Excellent; $0.7 \leq \alpha < 0.9$ Good; $0.6 \leq \alpha < 0.7$ Acceptable; $0.5 \leq \alpha < 0.6$ Poor; $\alpha < 0.5$ Unacceptable. Further, Cohen’s (1992) classification of Effect Sizes (ES) i.e. small (0.10); medium (0.30); large (0.50) and based upon Rosenthal’s (1996) research Very large (0.70) were used to determine the thresholds of $r$. These views are congruent with Field’s (2013, P. 82) assertions that $r$ "is an effect size" and associated, although not linearly, to Cohen’s (1992) classification of effect sizes.

A regression analysis was considered as part of the investigation. However, given the $r$ values for results relating to Pearson’s Correlation Coefficient were less than $r= 0.50$ (with the exception of the Abdominal Flexion test and the T-test) this approach was subsequently discarded as the results could not be determined as accurate or reliable.

**RESULTS**

Modifications made to the dataset, relating to normalising the data for linearity and homogeneity, resulted in an analysis of 18 active college aged students (males: n= 13, females: n= 5). The mean (± SD) for 13 males was 19.2 years ± 3.22 years and for 5 females was 19.4 years ± 1.14 years. Participants were also asked a series of questions, immediately prior to the test period, that included assessing: the number of days spent training in the participants sport specific area per week, 2.72 days ± 0.67 days; "the number of days spent training in the participants non-sport specific area per week", 1.94 ± 1.16 days; as well as "food consumption", 46.66 ± 34.47 min and "liquid consumption", 40 ± 23.00 minutes, prior to commencement of the test period.

The temperature (and relative humidity) across the three test days was recorded at 28ºC (30%), 32ºC (25%), and 28ºC (45%) for days 1, 2 and 3, respectively. Temperature and relative humidity conducted during the ICC measures relating to test-retest reliability was 32ºC and 24%.

A large correlation existed between the Abdominal Flexion test and the T-test, $r = .604$, n = 18 p= 0.008. Further, the side balance test ($r$) and the postero-medial results of the right leg on the star excursion balance...
test demonstrated a medium ES of \( r = 0.387, n = 18, p= 0.113 \).

Abdominal flexion test results and results for the anterior of the right leg on the star excursion balance test demonstrated a correlation of \( r = 0.435, n = 18, p= 0.071 \), equating to a medium \( r \). Other than this, no significant correlation was determined between the results of the remaining three core stability tests and the star excursion balance test.

Whilst the purpose of this paper was to investigate the correlations that may be evident between core stability assessments and dynamic balance and agility, other correlations of note were between the back extension test and abdominal flexion test \( (r = -0.84, n =18, p = .739) \). A large correlation \( (r=0.588, n =18, p=0.010) \) was also found between the Side balance test (Left) and Side balance test (Right). Both these correlations lacked significance.

A large correlation was found between the postero-medial right and left leg in the star excursion balance test \( (r = 517, n = 18, p= 0.013) \); very large between the anterior right and left leg in the Star excursion test \( r = 713, n = 18, p= 0.001 \) and large in the postero-lateral right and left leg in the Star excursion test \( r = 677, n = 18, p= 0.002 \).

Cronbach's ICC measures within session one for inter-rater reliability for items tested revealed an ICC of \( a = .722 \) (95% CI: .497 to .878) as per Field's (2013) stipulations. The internal consistency was determined as good.

**DISCUSSION**

The purpose of this study was to determine if tests used to assess core stability were a predictor of success in tests used to determine agility and dynamic balance. The basis of this hypothesis was that previous research in the area had determined links between the aforementioned components of fitness i.e. core stability and dynamic balance and core stability and agility. For instance, Granacher et al (2014) had indicated the role core stability plays in preventing slips, trips and falls. Further, Araujo et al's (2015), Lee's (2005) and Petersen and Nittinger’s (2014) research indicated that core stability aids dynamic movement through the principle of kinetic linkage that involves stabilisation of the core muscles.

Contrary to the hypothesis, the results of the current study indicate that there was no significant relationship between tests that assessed core stability and tests conducted to assess dynamic balance in active, young adults. Further, with the exception of the abdominal flexion test, no significant relationship existed between the remaining core stability tests and the agility T-Test.

Waldhelm and Li’s (2012) research aimed to determine the most reliable battery of tests to assess core stability. Thus, the abdominal flexion test; side balance test (left and right) and back extension tests were compared against a battery of core-related assessments that also reviewed strength, flexibility, motor control and functional tests. Evaluation of the tests was established using coefficient of variance as part of intra rater reliability tests. This was not facilitated in this study as a result of injuries sustained to some of the participants whilst practising their own sports that prevented accurate test retest procedures to take place. In a further distinction this study relied upon two assessors for each core assessment tested and three for the Star excursion balance test, in contrast one assessor was used in Waldhelm and Li’s (2012) study.

The Higher ICC results in the Star Excursion Balance Test in this study, compared with the core stability
tests, may be attributed to the use of three assessors. Disparity exists between the measures of reliability used in the aforementioned research where a test re-test reliability coefficient was applied in comparison with this paper’s method i.e. ICC measures relating to inter-test reliability. The inter-rater reliability for the abdominal flexion test, the back extension test and side balance tests (Left and Right) for this study were 95% CI: .707; .699; .701 and .704.

When assessing reliability i.e. the degree to which an assessment tool produces stable and consistent results, the studies presented as evidence offer high levels of test retest reliability e.g. Waldhelm and Li (2012) for the core stability tests; the T-Test (Raya et al, 2013) and the Star Excursion Balance Test (Gribble, et al, 2013).

Consistency between ICC measures in this study and Waldhelm and Li (2012), Raya et al (2013) and Gribble, et al (2013) indicate the tests, as independent measures of a component of fitness, are reliable. Variation between the results of these tests when correlated with one another may be an indication that they have lowered ecological validity.

Further to the above paragraph, whilst the function of this study was to investigate the relationship between core stability tests as a determinant of balance and agility, it is worth noting that, within individual tests, large correlations existed. These findings are supported by good inter-rater reliability scores i.e. ICC scores of 95% CI: .712; .720; .719 and .718, respectively. This supports the reliability of the test as a measure of dynamic balance, having normalised the results based upon Coughlan et al’s (2014) protocol.

Based upon Moreland, Finch, Stratford, Balsor and Gill’s (1997), Flint (1958) and Youdas, Garrett, Egan and Therneau’s (2000) research, strong correlations should have been evident between the abdominal flexion test and the back extension test given the relationship between abdominal endurance and lower back endurance. A very large negative correlation associated with the back extension test and abdominal flexion test was evident (r = -0.84, n =18, p = .739) This very large negative correlation may be as a consequence of fatigue influenced by successive core stability exercises that were conducted as a battery of tests together and, as previously identified, as part of a random sequence. It was intended that the random sequencing of assessments, within the core stability tests, would have reduced the influence of pre and post exhaustion between the anterior, posterior and lateral musculature of the core stabilising muscles.

A large correlation existed between the Abdominal Flexion test and the T-test (r = .604, n = 18, p= 0.008), which was further supported by inter-test ICC scores that were marginally outside Kline’s scale of classification to be determined as good i.e. a = .699. These findings indicate that the core stabilising musculature of the anterior abdominal area may have an effect upon performance in the agility T-Test. Further, there was a large correlation (r=.588, n 18, p=.010) found in the current study between the Side balance test (Left) and Side balance test (Right). However, this correlation was not significant possibly owing to the small sample size.

It was not within the remit of this study to investigate the validity of the tests used to assess core stability, agility and dynamic balance in active young adults; however this may have impacted upon the results obtained. A drawback of determining correlation between the above tests is that the tests themselves are limited in terms of their validity. For example, Waldhelm and Li’s (2012) study investigated the reliability of the core stability tests but not their validity in relation to sporting performance. Additionally, Pauole et al’s (2000) study regarding the use of the agility T-Test was conducted among a sample that demonstrated variation in levels and modes of sport participation. This was a factor that was also evident in this study where
participants who conducted the tests were drawn from a variety of sporting backgrounds. Furthermore, in this study, anthropometric measurements were not collated. However, the links between confounding factors such as Body Mass Index (BMI) and height and weight as determinants of performance in relation to core stability, agility or dynamic balance is not the subject of investigation in this research.

CONCLUSIONS

With the exception of abdominal flexion test and the correlation with the T-Test, there were low correlations between the measures of core stability, dynamic balance and agility. Therefore, it must be concluded that core stability does not influence performance in tests applied to assess dynamic balance and agility. These results could indicate that athletes do not require good core stability in order to have good balance and/or agility. The low ecological validity of the tests chosen to assess core stability, agility and balance is a confounding factor that has impacted upon the results of this study.

The importance of this finding for sports coaches and exercise instructors is that success in one area of the components of fitness and associated testing techniques may not be a valid predictor of success in another. More importantly, it is clear that sports and exercise coaches/instructors must ensure that the methods used to improve the components of fitness such as core stability, agility or dynamic balance are addressed individually when composing training regimes. Ultimately, these components of fitness must be viewed separately given that no correlation can be established between the validity of the tests to assess these components in relation to one another.

Further areas for research that would illuminate this topic area include investigating the impact of core stability training upon the outcomes of the battery of core stability tests outlined in this paper. Additionally, researchers may benefit from differentiating exercises that impact upon core stability and those that affect core strength and further determining which exercise regimes, if any, have a greater impact upon these two disparate components of fitness.

REFERENCES


34. Waldhelm, A., & Li, L. (2012). Endurance tests are the most reliable core stability related measurements. *Journal of Sport and Health Science, 1*(2), 121-128.
