Scientific Research

Concurrent and Criterion-Referenced Validity of Trunk Muscular Fitness Tests in School-Aged Children

Ryan D. Burns[1], James C. Hannon[1], Pedro F. Saint-Maurice[2], Gregory J. Welk[2]

[1]Department of Exercise and Sport Science, University of Utah, Salt Lake City, USA
[2]Department of Kinesiology, Iowa State University, Ames, USA
Email: ryan.d.burns@utah.edu

Abstract

A cause of limited physical activity levels in youth is the presence of low back pain (LBP), therefore proper assessment of low back function in physical education settings is needed to identify children who may be at risk. The purpose of this study was to determine the concurrent and criterion-referenced validity of field tests of low back and core muscular endurance in school-aged children. The sample consisted of 4th through 10th grade students (N = 370) who completed low back and core muscular fitness tests on four separate testing days during their physical education classes. Field measures related to low back function included the Box 90° Trunk Extension (Box 90°) and the FITNESSGRAM Trunk Extension (FG-TE). Field measures related to overall core function consisted of a Lateral Plank, Prone Plank, and a Static and Dynamic Curl-up. Criterion measures of low-back muscular endurance included the Parallel Roman Chair Dynamic Trunk Extension (PRC-DTE) and the Parallel Roman Chair Static Trunk Extension (PRC-STE). Multivariate analysis using canonical correlations showed moderate correlations between low back and core measures ($P < .001$). The Lateral Plank, Prone Plank, and Dynamic Curl-up had moderate-to-strong canonical cross-loadings with the low back measure variate. The FG-TE displayed an insignificant canonical coefficient, and weak canonical loadings and cross-loadings. Measures of overall core function also significantly agreed with the criterion measures in classifying students into ranked tertile groups ($P < .001$). These results suggest that assessment of specific low back muscular function can be easily evaluated using tests of overall core muscular endurance as an alternative to the FG-TE in physical education settings.

Keywords

Assessment, Core Function, FITNESSGRAM

1. Introduction

Low back pain (LBP) is the most common cause of disability among individuals under the age of 45 (Calvo-Munoz, Gomez-Conesa, & Sanchez-Meca, 2013; Cunningham & Kelsey, 1984) with an estimated annual cost of 20 to 50 billion from injury compensation, medical treatment, and lost workdays (Dagenais, Caro, & Halderman, 2008; Oliver, Adams-Blair, & Dougherty, 2010). LBP has been shown to commence during childhood or early adolescence (Feldman, Shrier, Rossingol, & Abenhaim, 2001) and some evidence suggests that back pain in youth has a high predictive value for back pain in adulthood (Brattberg, 2004; Leboeuf-Yde, Latimer, Maher, Refshauge, & Colaco, 1999; Salminen, Erkintalo, Laien, & Pentti, 1995). A frequent cause of decreased physical activity levelsin youth and adults is the presence of LBP (Walters, 2000). Indeed, proposed correlates of LBP in youth include poor physical fitness, decreased physical activity, poor hamstring flexibility, increased sedentary behaviors including TV and computer time, and rising overweight and obesity rates (Fairbank, Pynsent, Van Poortvliet, & Philips, 1984; Mierau, Cassidy, & Yong-Hing, 1989; Sjolie, 2002; Sjolie, 2004).

Research has shown a relationship between back extensor muscular endurance and LBP (Alaranta, Luoto, Heliovaara, & Hurri, 1995; Auvinen, Tammelin, Taimela, Zitting, & Karppinen, 2008; Biering-Sorenson, 1984). Biering-Sorenson (1984) completed extensive physical examinations of lumbar strength and range of motion, which was followed up by a low back injury questionnaire administered a year later. There was a direct correlation found between the presence of LBP and back extensor weakness in the sample of school-aged children. A subsequent study found that individuals who had poor performance on the Biering-Sorenson muscular endurance test were three times more likely to have LBP than those who demonstrated greater performance (Alaranta, Luoto, Heliovaara, & Hurri, 1995). Other studies have supported these findings regarding trunk extensor muscular endurance and LBP in the adult population (Jorgenson & Nicolaisen, 1987; Latimer, Maher, Refshauge, & Colaco, 1999; Nicolaisen & Jorgensen, 1985). In the youth population, Mierau, Cassidy and Yong-Hing (1989) found an association between hamstring flexibility and incidence of LBP and Olubusola, Chidozie, Christopher and Oyinade (2009) reported that decreased isometric back extensor endurance was associated with LBP. Other research has supported a link between back extensor muscular endurance and LBP in youth (Balague, Troussier, & Salminen, 1999; Newcomer & Sinaki, 1996).

To reduce the incidence of LBP in youth it is important to identify feasible and accurate field tests that can be administered in physical education classes to help identify students at risk. Hannibal, Plowman, Looney and Bradenburg (2006) examined the reliability and validity of the Box 90˚ dynamic trunk extension (Box 90˚) and the FITNESSGRAM Trunk Extension (FG-TE) field tests against criterion measures of low back dynamic and static muscular endurance, which consisted of the Parallel Roman Chair Dynamic Trunk Extension (PRC-DTE) and the Parallel Roman Chair Static Trunk Extension (PRC-STE). The Parallel Roman Chair tests were chosen as criterion measures because of their ability to isolate the erector spinae muscles of the low back similarly to the MedX dynamometer (Kearns, Brechue, Bauer, Pollock, & Fulton, 1997). The MedX can be used as a criterion measure, but because of its cost, time-consuming nature for proper assessment, and availability, the Parallel Roman Chair tests are suitable alternatives for research purposes. The field tests and criterion measures used in the Hannibal study demonstrated acceptable reliability, however only the Box 90˚ field test displayed strong correlation coefficients with the PRC-DTE, giving evidence of criterion-related (concurrent) validity for this field test in boys only. The FG-TE, a trunk extension test administered in many physical education classes, did not significantly correlate with any of the criterion measures. The Hannibal et al. (2006) study provided evidence for an alternative field test to the FG-TE for use in physical education trunk extensor assessment. However, the sample used in this particular study was limited to 40 adolescent males and 32 adolescent females and examined only one possible alternative to the FG-TE in the Box 90˚ field test.

Although shown to be a valid measure of trunk extensor muscular endurance, the Box 90˚ and other tests of low back muscular endurance (Roman Chair measures) are very time consuming and cumbersome to administer in physical education classes due to the use of specialized equipment that can only be used by a small number of students at any given time. Therefore other measures that can effectively be used to evaluate back extensor muscular endurance should be examined as possible alternatives. Measures of overall core muscular endurance have been considered to be potential indicators of LBP (Akuthota & Nadler, 2004; Sjolie, 2004) so they may be useful alternative (or complementary) field assessments.

Additional field tests that relate to core function and can feasibly be administered in physical education settings include the Lateral (Side Bridge) and Prone Plank tests, FITNESSGRAM's Dynamic Curl-up, and its iso-

kinetic analog the Static Curl-up. These field tests were not developed to specifically measure muscular endurance of trunk extensors (Meredith & Welk, 2010), however, they may have predictive value in determining LBP risk in the pediatric population. The direct evaluation of these tests with more accurate measures of trunk extensor muscles can provide additional insights about their utility in diagnosing youth at risk for LBP. Therefore, the purpose of this study was to determine the concurrent and criterion-referenced validity of low back and core muscular fitness tests against criterion measures in a sample of 4th - 10th grade students. It was hypothesized that the majority of the field tests examined would have strong relationships and agreement with the criterion measures, indicating that both specific low back movements, in addition to overall core movements, can be effective assessments of low back muscular endurance in physical education settings.

2. Methods

2.1. Participants

The sample consisted of 370, fourth through tenth grade students (198 boys, 172 girls) recruited from three private schools located in a metropolitan area in the southwestern US. The sample included 61 fourth graders, 56 fifth graders, 56 sixth graders, 49 seventh graders, 48 eighth graders, 50 ninth graders, and 50 tenth graders. All students were free from injury or any condition that may have precluded them from participating in physical activity. Written consent was obtained from parents and written assent was obtained from the participants prior to data collection. The University Institutional Review Board and principals from the participating schools approved the protocols used in this study.

2.2. Criterion Tests

The *PRC-DTE* required students on a Roman Chair with their legs secured by a partner. Participants started at a 180-degree angle between back and legs while keeping their arms folded across the chest (neutral position). The students flexed their trunk until they reached a 90-degree angle between trunk and legs and then extended back to the neutral position. A chain attached from the students' shirt indicated when the participant had reached the 90-degree threshold of the repetition (when the attached paperclip on the chain barely touched the floor at the 90-degree flexion angle). Participants performed the trunk extension lifts at a specific cadence of 1 repetition every 3 seconds. The test was terminated when the student was unable to keep with the 20 contractions per minute cadence or voluntarily decided to stop. The final number of consecutive repetitions was used for analysis (Hannibal et al., 2006).

The *PRC-STE* required the students simply holding a neutral position (180 degrees) with arms crossed as long as they could on a Roman Chair apparatus. The test was terminated if either an attached paper clip (for reference) touched the floor during 2 or more consecutive seconds or a hyperextension beyond 180 degrees occurred (8). Total time in the neutral position was recorded for analysis.

2.3. Low Back Muscular Fitness Tests

The *Box* 90˚ test consisted of students lying over a 1/2 inch foam pad that was placed over a 42" × 36" × 36" wooden box. Participants had to flex forward until a 90-degree angle was reached (upper body is parallel to the box) and then extend back to a neutral position (180 degrees at the starting position). Each contraction was set at a 3 second cadence and the test was terminated if the participant is not able to keep this cadence for two consecutive times or voluntarily wished to stop.

The *FG-TE* test consisted of having the participants lie in the prone position with hands under the thighs and lifting their trunk off the floor as high as possible. This upward movement was to be performed in a controlled manner keeping the head aligned with the spine. Once the highest trunk elevation had been achieved, the participant was to hold this position for 2 - 5 seconds with the distance between the chin and the floor measured by the observer with a standard ruler. The feet had to be in contact with the floor during the test. Distance from the chin to the floor was recorded in inches and used for analysis (Meredith & Welk, 2010).

2.4. Overall Core Fitness Tests

FITNESSGRAM's *Dynamic Curl-up* consisted of having the students curl-up and down sliding their fingers

across a distance of 3 inches (5 - 9 years old) or 4.5 inches (older students) at a specific cadence provided by a recorded CD. On each curl-up, participants had to touch with their back and head on a mat. Only 2 errors were allowed before the test was terminated. These procedures are in accordance to those recommended by FITNESSGRAM (Meredith & Welk, 2010).

The *PronePlank test* consisted of having the students maintain a 90-degree angle between elbows and the trunk. Only the elbows and toes were allowed to be in contact with the mat and any corrections that had to be made were in a 3 second period otherwise the test was terminated. Total time recorded maintaining the proper position was used for analysis.

The *Lateral Plank* test required the students to hold his/her body in a straight line, perpendicular to his/her right elbow (90 degrees) with only elbows and shoes in contact with the mat for the longest time possible. Any correction made during the exercise had to be adopted within 3 seconds or the test was terminated. The recorded time in the proper position was recorded for analysis.

Finally, the *Static Curl-up* was a modified version of the FITNESSGRAM Curl-up test. The same principles applied from the Dynamic Curl-up test although participants had to maintain the "up" position (body at 45-degree angle, arms extended) for as long as possible with the fingers staying at the end of the 4.5-inch wide stripe. If the fingers moved from the edge of the stripe they were instructed to go back to the correct position within 3 seconds or the test was terminated. The total time maintaining an appropriate position was used for analysis.

2.5. Procedures

All data were collected during the students' regular physical education classes. The students were involved in a larger evaluation of youth fitness profiles and practiced each of the planned tests on multiple occasions prior to data collection to gain familiarity with the assessments. Students completed the assessments of low back and core muscular fitness on 4 separate testing days with 1 week separating consecutive testing sessions. Trained graduate students within the Department of Exercise and Sport Science administered all laboratory and field tests. Each testing day consisted of 2 - 3 stations where the students completed the various fitness tests scheduled for that respective day. Students were randomly assigned to a station at the beginning of their physical education class and then rotated to the next station when an assessment was successfully completed. At least 5 minutes were given between consecutive tests of low back and core muscular fitness. The first day of testing included two testing stations to administer the Dynamic Curl-up and the PRC-STE. The second day consisted of three stations where the Prone Plank, Lateral Plank, and PRC-DTE assessments were administered. The third day also used three testing stations where the Box 90˚, Static Curl-up, and FG-TE tests were administered. The fourth day was used as a make-up day for students to complete tests that were not completed on the other days.

2.6. Statistical Analysis

Data was initially screened for outliers and checked for normal distributions using k-density plots and the Shapiro-Wilk test on all descriptive and fitness test measures. Differences between the genders and among grade levels on the descriptive data were analyzed using multiple 2×3 Factorial ANOVA tests. Age groups were stratified into 3 grade levels: elementary (fourth and fifth grade), middle school (sixth through eighth grade), and high school (ninth and tenth grade). If significant differences were found among the grade levels, a Bonferroni post hoc analysis was employed. Alpha level was adjusted appropriately using the Bonferroni adjustment.

Concurrent validity was examined between criterion and field measures employing Pearson correlations using the total sample for analysis. Correlations were considered strong if $r \geq .60$, moderate if $r = .30$ to $.60$, and weak if $r \leq .30$ (Pagano & Gavreau, 2000). Multivariate analysis was conducted using canonical correlations. Canonical correlations maximize the association between two separate weighted linear combination of variables. A weighted linear combination of variables is referred to as a variate. Two variates were examined in this study: one consisting of measures of low back function (PRC-DTE, PRC-STE, Box 90˚, and FG-TE) and one variate consisting of measures of overall core muscular endurance (Dynamic Curl-up, Static Curl-up, Prone Plank, and Lateral Plank). This approach was employed to examine several phenomena including: if low back measures significantly correlate with overall core measures within a multivariate framework, which individual measures contribute the most and correlate the strongest with their respective variate, and finally, which individual measures correlate the strongest with the opposite variate. Canonical correlation analysis may yield several significant

functions, however only the functions that explain a significant portion of shared variance between the two variates will be analyzed. Which canonical functions will be analyzed will be based off of the Redundancy Index (Rd) for each significant function. Based on the practical significance of the Rd, the raw and standardized coefficients, canonical loadings (the correlation between an individual measure and its variate), and the canonical cross-loadings (the correlation between an individual measure and the opposite variate) will be reported.

Criterion-referenced validity was examined using group classification of fitness score data. Muscular fitness scores were stratified into tertiles of equal number based on ranked scores within each grade-gender group. Each student's muscular fitness test score (from criterion and field measures) were recoded into a three-group classification scheme representing the lower, middle, and upper ranked tertile group. This approach was employed to gather preliminary evidence on criterion-referenced validity and therefore, determine the extent criterion and field measures could agree if health-risk cut-points were developed in future research studies. This is particularly important in pediatric research since fitness scores are commonly used to allocate youth into different "zones" (e.g. FITNESSGRAM Healthy Fitness Zones) using established cut-points related to health outcomes. Criterion-referenced agreement between criterion and field measures into these tertile groups was statistically analyzed using weighted kappa statistics and percentage of agreement. Alpha level was set at $P \leq .05$ and all analyses were carried out using STATA v12.0 (College Station, TX, USA) statistical software.

3. Results

3.1. Descriptive Differences

Means and standards deviations for the descriptive characteristics in each grade and gender group are presented in **Table 1**. In this sample, age, height, weight, and BMI were statistically higher at the older grade levels ($P < .01$). Boys were also taller and weighed more than girls ($P < .05$), but there were no gender differences in BMI.

3.2. Pearson Correlations

Table 2 presents the Pearson correlations between criterion and field measures. When examining the relationships including the total sample for analysis, the Box 90° field test had the strongest correlation coefficient with any criterion measure ($r = .51$, $P < .001$) (**Figure 1**). The Lateral Plank and Prone Plank also had statistically significant moderate correlations with PRC-DTE ($P < .05$). The FG-TE had relatively weak correlations with the criterion measures compared to the other field tests.

3.3. Canonical Correlations

Canonical correlation analysis yielded two statistically significant functions. The first canonical function yielded a correlation of .477 (Wilk's $\Lambda = .730$, $F_{(16, 1125)} = 7.65$, $P < .001$), explaining 22.7% of variance between variates; the second canonical function yielded a correlation of .201 (Wilk's $\Lambda = .945$, $F_{(9, 898)} = 2.35$, $P < .05$), explaining 4.04% of variance. The Rd for the first canonical function was 11.5% and the Rd for the second function was .3%. Because the first canonical function explained significantly more variance between the two variates, and yielded a higher Rd, it was the only function that was further analyzed. **Table 3** depicts the raw and standardized coefficients, canonical loadings, and canonical cross-loadings for the first canonical function. The Box 90° and PRC-DTE had the strongest standardized weights for the low back variate, the strongest correlations with the low back variate (loadings), and the strongest correlation with the core variate (cross-loadings). The FG-TE's coefficient was not statistically significant nor did it strongly correlate with any of the variates. The Lateral Plank and Dynamic Curl-up had the strongest coefficients from the core variate. The Lateral Plank, Prone Plank, and Dynamic Curl-up all displayed moderate to strong correlations with the core variate and with and low back variate. The Static Curl-up's correlations were relatively weak compared to the other measures.

3.4. Criterion-Referenced Validity

Table 4 displays the weighted kappa statistics and percentage of agreement into tertile groups between criterion and field measures. Agreement ranged from 62.2% to 87.7% between the muscular fitness test scores and criterion measures. The Box 90° yielded the strongest weighted kappa statistics and agreement among all field tests examined. The FG-TE did not significantly agree with any of the criterion measures.

Table 1. Descriptive data by grade and gender group.

	Elementary (n = 116)	Middle School (n = 154)	High School (n = 100)	Girls (n = 172)	Boys (n = 198)
Age (years)	9.96 ± .71	**12.43 ± .90**[*]	**15.19 ± .66**[**]	12.25 ± 2.17	12.51 ± 2.10
Height (m)	1.41 ± .08	**1.56 ± .10**[*]	**1.68 ± .10**[**]	1.53 ± .12	**1.56 ± .15**[†]
Weight (kg)	34.87 ± 8.00	**48.32 ± 12.45**[*]	**60.77 ± 12.40**[**]	45.63 ± 13.72	**49.00 ± 15.75**[†]
BMI (kg/m^2)[1]	17.40 ± 2.91	**19.42 ± 3.57**[*]	**21.35 ± 3.43**[**]	19.07 ± 3.73	19.51 ± 3.58

Note: [1]BMI Stands for Body Mass Index; [*]Significantly higher score than Elementary, $P < .01$; [**]Significantly higher score than Elementary and Middle/High School, $P < .01$; [†]Significant gender differences, $P < .05$.

Table 2. Zero-order Pearson correlations between criterion and field measures.

Criterion Measures	Field Measures					
	Box 90°[3]	FG-TE[4]	Lateral Plank	Prone Plank	Static Curl-up	Dynamic Curl-up
PRC-DTE[1]	**.51**[†]	**.15**[†]	**.42**[†]	**.43**[†]	**.16**[†]	**.35**[†]
PRC-STE[2]	**.30**[†]	**.21**[†]	**.18**[†]	**.18**[†]	**.22**[†]	**.22**[†]

Note: [1]PRC-DTE stands for Parallel Roman Chair Dynamic Trunk Extension; [2]PRC-STE stands for Parallel Roman Chair Static Trunk Extension; [3]Box 90° stands for BOX 90-degree dynamic trunk extension; [4]FG-TE stands FITNESSGRAM Trunk Extension; [†]Statistically significant, $P < .05$.

Table 3. Canonical coefficients for the first canonical function.

	Raw Coefficient	Significance	Standardized Coefficient	Canonical Loading	Canonical Cross-loading
Low Back Variate					
PRC-DTE[1]	**.023**[†]	.000	.447	.767	.366
PRC-STE[2]	**.001**[†]	.038	.205	.399	.191
Box 90°[3]	**.034**[†]	.000	.639	.886	.423
FG-TE[4]	.081	.512	.064	.131	.063
Core Variate					
Prone Plank	**.002**[†]	.050	.240	.742	.354
Lateral Plank	**.010**[†]	.000	.493	.863	.412
Dynamic Curl-up	**.018**[†]	.000	.403	.741	.354
Static Curl-up	**.001**[†]	.010	.252	.384	.183

Note: [1]PRC-DTE stands for Parallel Roman Chair Dynamic Trunk Extension; [2]PRC-STE stands for Parallel Roman Chair Static Trunk Extension; [3]Box 90° stands for BOX 90-degree dynamic trunk extension; [4]FG-TE stands FITNESSGRAM Trunk Extension; [†]Statistically significant, $P < .05$.

Table 4. Tertile agreement between criterion and field measures (reported as Weighted Kappa and Percentage of Agreement).

Field Measures	Criterion Measures			
	PRC-DTE[1]		PRC-STE[2]	
	Kappa	Agreement	Kappa	Agreement
Box 90[3]	**.421**[†]	**87.9%**	**.267**[†]	**81.5%**
FG-TE[4]	.011	67.4%	.050	62.2%
Lateral Plank	**.334**[†]	**85.7%**	**.179**[†]	**78.2%**
Prone Plank	**.276**[†]	**84.3%**	**.134**[†]	**76.5%**
Static Curl-up	.040	73.3%	**.136**[†]	**77.2%**
Dynamic Curl-up	**.268**[†]	**81.3%**	**.138**[†]	**78.6%**

Note: [1]PRC-DTEstands for Parallel Roman Chair Dynamic Trunk Extension; [2]PRC-STEstands for Parallel Roman Chair Static Trunk Extension; [3]Box 90stands for BOX 90° dynamic trunk extension; [4]FG-TEstands for FITNESSGRAM Trunk Extension; [†]Statistically significant, $P < .05$.

4. Discussion

The purpose of this study was to examine the concurrent and criterion-referenced validity of field measures of low back function and core muscular endurance in a sample of school-aged children. The concurrent validity
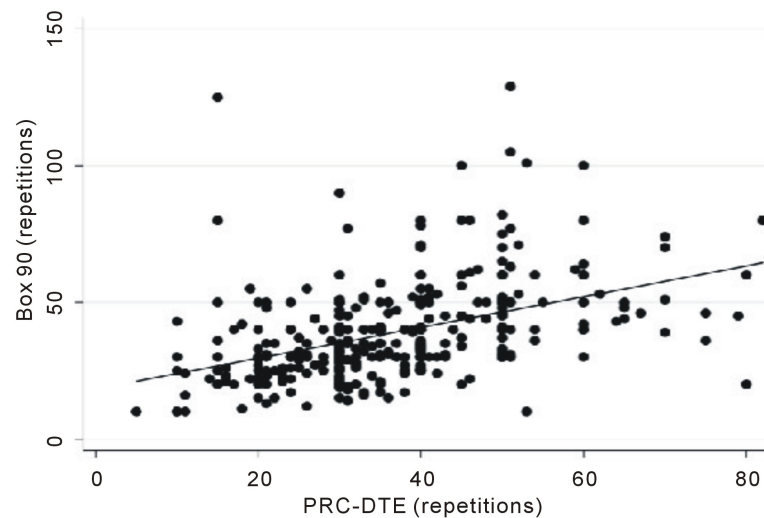
Figure 1. Scatter plot and line of best fit displaying the relationship between PRC-DTE scores and Box 90° scores for the total sample (N = 370).

evidence when evaluating the low back field tests against the criterion measures showed that the strongest Pearson correlation coefficients were seen between the Box 90° and PRC-DTE. The FG-TE significantly correlated with PRC-DTE and PRC-STE but the magnitude of correlations were weak compared to the other field measures. Other field measures such as the Lateral Plank and Prone Plank had moderate correlations with the criterion tests.

Congruent with previous research, it was expected that the Box 90° field test would produce stronger correlations with the PRC-DTE due to the similarity of movement between the two tests and the shared dynamic muscular endurance component of fitness. Although these two movements are very similar, the correlations between the Box 90° and PRC-DTE were only considered moderate. One possible explanation for this is that to help stabilize the participants' lower body during the Box 90°, two students had to hold the participants' legs throughout the duration of the test. This may have decreased the stabilization of the lower body compared to the PRC-DTE, where the students' legs were supported by the apparatus, and could have had a significant effect on the results. Despite this however, the Box 90° consistently had significant correlations with the criterion measures.

The incorporation of the Box 90° into physical education classes may be beneficial in terms of administering a trunk extension test that would provide more useful information regarding low back muscular endurance compared to the FG-TE. Roman Chairs, as suggested by Hannibal et al. (2006), are not prohibitively expensive for many school programs as it is a common piece of equipment found in many weight rooms. However, the Box 90° or a derivative of it, perhaps a bench or table in replacement of a wooden box, may save many programs money in trying to incorporate a more valid assessment of low back muscular strength and endurance into their programs. Despite the benefits of using the Box 90° for back extensor assessment, a limitation to incorporating this movement in physical education settings is that it is a time-consuming test to administer to large class sizes. Testing time using the Box 90° and PRC tests in this study took, on average, approximately 3 minutes per student. If limited equipment is available for the physical educator to administer these assessments, testing an entire class can easily take up an entire class period (or more).

The FG-TE, used in FITNESSGRAM's fitness assessment battery, was the primary field test examined in this study for comparison with the Box 90°. FG-TE is considered a test of trunk extensor strength and flexibility, albeit with limited back extensor strength needed during the short 3 - 5 sec contraction (Pollock et al., 1989). The weak correlations between FG-TE and the criterion measures were expected as the Roman Chair tests involved low back muscular endurance rather than lumbar flexibility. The FG-TE is used in the FITNESSGRAM fitness assessment battery as a test of low back strength and flexibility (Meredith & Welk, 2010). However, results from this study, in support of Hannibal et al. (2006), show that the FG-TE did not significantly relate to other tests of low back muscular endurance, nor did it significantly relate to tests of core muscular endurance. Indeed, from the results of the multivariate canonical correlation analysis, the FG-TE was the only measure not to have a

statistically significant coefficient. These results indicate that the FG-TE should not be used as a measure of low back muscular endurance, thus also should not be used as a proxy measure for LBP risk assessment. Rather, it should be used only as a measure of lumbar flexibility (as opposed to lumbar strength) as specified by Meredith & Welk (2010).

The other muscular fitness tests examined in this study were not exclusive to low back muscle activation, but also involved the gluteal, hamstrings, abdominals, and in the case of the Prone and Lateral Plank tests the use of the muscles of the upper body as well (Hibbs, Thompson, French, Hodgson, & Spears, 2011). Despite this, their utility in physical education classes should not be overlooked as valid measures of back extensor function as the Lateral and Prone Plank, in addition to the Dynamic Curl-up, showed statistically significant correlations with the low back and core variates, along with strong classification agreement with the criterion measures with values at approximately 80% (see **Table 3**). These results provide preliminary evidence of the concurrent and criterion-referenced validity of the overall core measures, suggesting that if health-related cut-points were to be developed for the muscular fitness test scores, the low back criterion and overall core measures would produce similar classification of students into a three Fitness Zone classification scheme. From a health-related perspective and assuming LBP is related to low back and core muscular endurance (Balague, Troussier, & Salminen, 1999; Behm & Anderson, 2006), this approach is useful since it can provide insights on the ability of field tests to classify children who are at risk for developing LBP based on testing performance. Incorporating these assessments into physical education will also increase the value of the lesson if physical educators communicate the importance of low back health while concomitantly allowing the students to perform movements that will strengthen the musculature of the low back.

There were a few limitations in this study that may have influenced the results. Pelvic stabilization is important for isolating the erector spinae muscles of the low back, eliminating potential significant contribution from the gluteal musculature and hamstrings during the movement (Graves, Webb, Pollock, Leggett, & Carpenter, 1990). Perhaps simply holding the legs during the Box 90˚ assessment was not effective in providing acceptable pelvic stabilization. Also, students completed consecutive low back and core assessments on a testing day. Although ample time was given between assessments (at least a 5-minute recovery), fatigue may have been an issue for some students, especially following the Roman Chair tests. Finally, this was a cross-sectional study merely displaying the relationships between low back and core muscular fitness testing performance. Therefore, no inferences can be made about the protective effect of low back and core muscular endurance on the incidence of LBP in children. Future research should focus on determining the relationships between fitness testing performance and back health using longitudinal research designs incorporating a valid assessment of the health of the low back.

This paper suggests that the FG-TE, a commonly used assessment in FITNESSGRAM for lumbar strength and flexibility, should not be used as a measure of low back muscular endurance in school-aged children. Since low back muscular endurance may be an indicator of low back health, the Box 90˚, Lateral Plank, Prone Plank, or Dynamic Curl-up tests may give physical educators more meaningful information than the FG-TE regarding the health of the low back. This study also supports that measures of overall core muscular endurance can be useful assessments to examine specific low back muscular endurance. Physical educators and specialists do not need to administer cumbersome tests such as the PRC-DTE or Box 90˚ to effectively evaluate back extensors in physical education settings. Additionally, combining these assessments with hamstring stretches at the beginning of a physical education lesson, prior to more vigorous movements, on a weekly basis enhances the value of physical education by allowing the student to learn about and engage in movements that may be beneficial to their long-term back health. Despite these findings, the question remains if these tests can be used to accurately classify students at risk for developing LBP. Future research needs to examine what specific tests or combination of tests show the strongest relationships with LBP prevalence and incidence. Doing this will enable researchers to develop more meaningful cutoff scores for field measures that can be used to classify students based on testing performance.

Acknowledgements

References

Akuthota, V., & Nadler, S. (2004). Core Strengthening. *Archives of Physical Medicine and Rehabilitation, 85,* S86-S92. http://dx.doi.org/10.1053/j.apmr.2003.12.005

Alaranta, H., Luoto, S., Heliovaara, M., & Hurri, H. (1995). Static Back Endurance and the Risk of Low Back Pain. *Clinical Biomechanics, 10,* 323-324. http://dx.doi.org/10.1016/0268-0033(95)00002-3

Auvinen, J., Tammelin, T., Taimela, S., Zitting, P., & Karppinen, J. (2008). Associations of Physical Activity and Inactivity with Low Back Pain in Adolescents. *Scandinavian Journal of Medicine and Science in Sport, 18,* 188-194. http://dx.doi.org/10.1111/j.1600-0838.2007.00672.x

Balague, F., Troussier, B., & Salminen, J. (1999). Non-Specific Low Back Pain in Children and Adolescents: Risk Factors. *European Spine Journal*, *8*, 429-438. http://dx.doi.org/10.1007/s005860050201

Behm, D. G., & Anderson, K. (2006). The Role of Instability with Resistance Training. *Journal of Strength and Conditioning Research, 20,* 716-722.

Biering-Sorenson, F. (1984). Physical Measurements as Risk Indicators for Low Back Trouble over a One-Year Period. *Spine, 9,* 106-119.

Brattberg, G. (2004). Do Pain Problems in Young School Children Persist into Early Adulthood? A 13-Year Follow-Up. *European Journal of Pain, 8,* 187-199. http://dx.doi.org/10.1016/j.ejpain.2003.08.001

Calvo-Munoz, I., Gomez-Conesa, A., & Sanchez-Meca, J. (2013). Prevalence of Low Back Pain in Children and Adolescents: A Meta-Analysis. *BMC Pediatrics, 13,* 1-12. http://dx.doi.org/10.1186/1471-2431-13-14

Cunningham, L., & Kelsey, J. (1984). Epidemiology of Musculoskeletal Impairments and Associated Disability. *American Journal of Public Health, 74,* 574-579. http://dx.doi.org/10.2105/AJPH.74.6.574

Dagenais, S., Caro, J., & Halderman, S. (2008). A System Review of Low Back Pain Cost of Illness Studies in United States and Internationally. *Spine Journal, 8,* 8-20. http://dx.doi.org/10.1016/j.spinee.2007.10.005

Fairbank, J., Pynsent, P., Van Poortvliet, J., & Phillips, H. (1984). Influence of Anthropometric Factors and Joint Laxity in the Incidence of Adolescent Back Pain. *Spine, 9,* 461-464. http://dx.doi.org/10.1097/00007632-198407000-00007

Feldman, D., Shrier, I., Rossignol, M., & Abenhaim, L. (2001). Risk Factors for the Development of Low Back Pain in Adolescence. *American Journal of Epidemiology, 154,* 30-36. http://dx.doi.org/10.1093/aje/154.1.30

Graves, J., Webb, D., Pollock, M., Leggett, S., & Carpenter, D. (1990). Pelvic Stabilization during Resistance Training: Its Effect on the Development of Lumbar Extension Strength. *Archives of Physical Medicine and Rehabilitation, 75,* 210-215.

Hannibal III, N. S., Plowman, S., Looney, M. A., & Bradenburg, J. (2006). Reliability and Validity of Low Back Strength/ Muscular Endurance Field Tests in Adolescents. *Journal of Physical Activity and Health, 3,* S78-S89.

Hibbs, A. E., Thompson, K. G., French, D. N., Hodgson, D., & Spears, I. R. (2011). Peak and Average Rectified EMG Measures: Which Method of Data Reduction Should Be Used for Assessing Core Training Exercises? *Journal of Electromyography and Kinesiology, 21,* 102-111. http://dx.doi.org/10.1016/j.jelekin.2010.06.001

Jorgenson, K., & Nicolaisen, T. (1987). Trunk Extensor Endurance: Determination and Relation to Low-Back Trouble. *Ergonomics, 30,* 259-267. http://dx.doi.org/10.1080/00140138708969704

Kearns, C. F., Brechue, W. F., Bauer, J., Pollock, M. L., & Fulton, M. (1997). Muscle Activiation during Isolated and Non-Isolated Lumbar Extension Exercise. *Medicine and Science in Sports and Exercise, 29,* S165. http://dx.doi.org/10.1097/00005768-199705001-00943

Latimer, J., Maher, C. G., Refshauge, K., & Colaco, I. (1999). The Reliability and Validity of the Beiring-Sorensen Test in Asymptomatic Subjects and Subjects Reporting Current or Previous Nonspecific Low Back Pain. *Spine, 24,* 2085-2090. http://dx.doi.org/10.1097/00007632-199910150-00004

Leboeuf-Yde, C., Latimer, J., Maher, C. G., Refshauge, K., & Colaco I. (1999). The Reliability and Validity of the Beiring-Sorensen Test in Asymptomatic Subjects and Subjects Reporting Current or Previous Nonspecific Low Back Pain. *Spine, 24,* 2085-2090. http://dx.doi.org/10.1097/00007632-199910150-00004

Meredith, M. D., & Welk, G. J. (2010). *Fitnessgram/Activitygram Test Administration Manual* (4th ed.). Champaign, IL: Human Kinetics.

Mierau, D., Cassidy, J., & Yong-Hing, K. (1989). Low-Back Pain and Straight Leg Raising in Children and Adolescents. *Spine, 14,* 526-528. http://dx.doi.org/10.1097/00007632-198905000-00010

Newcomer, K., & Sinaki, K. (1996). Low Back Pain and Its Relationship to Back Strength and Physical Activity in Children. *Acta Paediatrica, 85,* 1433-1439.

Nicolaisen, T., & Jorgensen, K. (1985). Trunk Strength, Back Muscle Endurance and Low-Back Trouble. *Scandinavian Journal of Rehabilitative Medicine, 17,* 121-127.

Oliver, G. D., Adams-Blair, H. R., & Dougherty, C. P. (2010). Implementation of a Core Stability Program for Elementary School Children. *Athletic Training and Sports Health Care, 2,* 261-266. http://dx.doi.org/10.3928/19425864-20100630-01

Olubusola, J. E., Chidozie, M. E., Christopher, A. O., & Oyinade, A. A. (2009). Isometric Endurance of the Back Extensors in School-Aged Adolescents with and without Low Back Pain. *Journal of Back and Musculoskeletal Rehabilitation, 22,* 205-212.

Pagano, M., & Gavreau, K. (2000). *Principles of Biostatistics* (2nd ed.). Belmont, CA: Brooks/Cole, Cengage Learning.

Pollock, M., Leggett, S., Graves, J., Jones, A., Fulton, M., & Cirulli, J. (1989). Effect of Resistance Training on Lumbar Extension Strength. *American Journal of Sports Medicine, 17,* 624-629. http://dx.doi.org/10.1177/036354658901700506

Salminen, J., Erkintalo, M., Laine, M., & Pentti, J. (1995). Low Back Pain in the Young: A Prospective Three-Year Follow-Up Study of Subjects with and without Low Back Pain. *Spine, 20,* 2101-2108. http://dx.doi.org/10.1097/00007632-199510000-00006

Sjolie, A. (2002). Psychosocial Correlates of Low Back Pain in Adolescents. *European Spine Journal, 11,* 582-588. http://dx.doi.org/10.1007/s00586-002-0412-z

Sjolie, A. (2004). Low Back Pain in Adolescents Is Associated with Poor Hip Mobility and High Body Mass Index. *Scandinavian Journal of Medicine and Science in Sports, 14,* 168-175. http://dx.doi.org/10.1111/j.1600-0838.2003.00334.x

Walters, P. H. (2000). Back to the Basics, Strengthening the Neglected Lower Back. *ACSM's Health and Fitness Journal, 4,* 19-25.