Pectoral and Hip Flexors Muscle Length Tests as Elements of Developmental Age Postural Examination Standardized for the Purpose of Interdisciplinary Utilization

Anna Gogola[1], Agnieszka Polaczek[1], Piotr Woźniak[2], Tomasz Wolny[1]

[1] Academy of Physical Education in Katowice, Poland
[2] University of Physical Education in Krakow, Poland

HOW TO CITE:
Gogola, A., Polaczek, A., Woźniak, P., & Wolny, T. (2025). Pectoral and hip flexors muscle length tests as elements of developmental age postural examination standardized for the purpose of interdisciplinary utilization. *International Journal of Special Education, 40*(1), 134-144.

CORRESPONDING AUTHOR:
Anna Gogola;
aniagogola@op.pl

DOI:
https://doi.org/10.52291/ijse.2025.40.11

ABSTRACT:

Proposed by the World Health Organisation, the bio-psycho-social model of medicine encourages specialists across various fields to embrace interdisciplinary approaches. For such teams to function effectively, members must acquire a foundational understanding of disciplines beyond their expertise. This facilitates the timely introduction of other essential interventions. The study's primary aim was to evaluate the reliability of the modified versions of clinical tests for the length of pectoral and hip flexor muscles. Additionally, the study sought to determine whether these modified tests could be utilized in scientific, home, or school settings, enabling non-clinical specialists to conduct screenings and identify the need for referral to a clinician.

Sixty-two children (36 girls) were enrolled in primary school classes four to eight, with a mean age of 12.15 years. The participants were generally healthy and displayed varying levels of daily physical activity. A repeated-measures design assessed the results of two muscle length tests conducted by two raters. Measurements were taken by both raters on the same day, as well as after two- and five-day intervals. Various reliability indices were calculated, demonstrating good to excellent reliability for the two muscle length tests.

Following appropriate, specific modifications, the two muscle length tests showed potential for achieving a high level of reliability, supporting the use of their modified versions in scientific settings. Additionally, other professionals – such as teachers, trainers, and educators – could benefit from employing these standardized and modified tests in their regular environments (e.g., home, school) for screening purposes and determining the need for referral to a clinician.

Keywords: Postural, Adolescents, Evaluation, Muscle.

INTRODUCTION

Proposed by the World Health Organization, the bio-psycho-social model of medicine requires specialists across various fields to adopt a comprehensive, interdisciplinary approach. This is especially relevant during the developmental stages of life when pathological processes can overlap with developmental changes, leading to significant dysfunctions that affect the entire body and various aspects of a patient's life, including their mental and social well-being. Therefore, members of interdisciplinary therapeutic teams need to gain foundational knowledge in areas outside their specialties. This enables them to conduct basic assessments and implement necessary interventions at an early stage. In this article, we aim to present two key components of postural examination for children and adolescents. These assessments can be utilized by various individuals in the child's environment, such as teachers, parents, and educators. Proper implementation and interpretation of these assessments can facilitate prompt referral to a physiotherapist or medical doctor, or enable monitoring of previously identified changes.

LITERATURE REVIEW

Other specialists should note that in the fields of physiotherapy, biomechanics, neurophysiology, and other disciplines related to human motor activity, there is an ongoing pursuit of new methods for data quantification. The technological advances observed in the twentieth and twenty-first centuries have significantly expanded the range of research tools available. Solutions employed in this area include motion analysis systems for assessing spinal mobility, pelvic configuration (Biały et al., 2010; Gnat & Biały, 2015; Gnat et al., 2015), gait, and other complex motor tasks (Hecht et al., 2022; Qiu et al., 2023), ultrasounds for examining the morphology and activity of various muscle groups (Chmielewska et al., 2023; Gnat et al., 2012), including the deep trunk muscles (Biały et al., 2017; Biały et al., 2019; Gogola et al., 2016; Gogola et al., 2018; Park, 2013), myotonometry (Kurashina et al., 2022; Kurashina et al., 2023), and elastography (Linek et al., 2021; Ormachea & Parker, 2020) for assessing the biomechanical properties of body tissues and their changes in response to various therapeutic techniques or forms of training (Dębski et al., 2019). Additionally, measurements of electrical muscle activity, such as electromyography in its multiple forms (Guruhan et al., 2021; Oliva-Lozano & Muyor, 2020) or functional magnetic resonance imaging of the brain during motor tasks (Gnat et al., 2021; Gnat et al., 2020), are also utilized. The precision required for these measurement techniques in scientific settings is gradually making its way into the clinical field and, in some cases, even patients' homes. This often comes at the cost of slightly reduced precision but offers the benefit of minimized time consumption and increased accessibility for patients.

One might also consider whether it would be both feasible and sensible to implement diagnostic techniques in the reverse direction, from the clinical field to the scientific one. The answer to the second part of this question is yes. Indeed, many specialists working in the 'no-man's land' between clinic and science, who seek to obtain scientific evidence to support the effectiveness of their preferred therapeutic techniques, often find that despite the abundance of technological instrumentation, it can unexpectedly prove insufficient. Despite their high level of sophistication, these devices may not capture the broader aspects of movement. Paradoxically, these advanced tools might be too precise, revealing a spectrum of real-world perception too narrow. In some cases, more basic measurements may be required to represent reality from a broader perspective. However, to apply used clinical tests in other settings, they must first undergo appropriate modification. This raises questions about the accuracy and reliability of the proposed diagnostic/measurement techniques.

In this particular study, we aimed to modify two widely used clinical muscle length tests and assess the reliability of their modified versions. The tests in question for the pectoral and hip flexor muscles are essential as these muscles can significantly affect body posture in children and adolescents. These tests form standard components of postural examinations during the developmental years. Shortenings of these muscles constitute characteristic features of the typical developmental postural syndromes described by Janda as the upper and lower crossed syndromes (D'Silva et al., 2023; Gilliani et al., 2020). The first of the tests is also known as the wall test, Dega test, or sitting arm rise test (Barczyk et al., 2009; Boguszewski et al., 2016), and the second – the Thomas test (Barczyk et al., 2009; Cady et al., 2022; Kim & Ha, 2015). This terminology is still in use in many clinical settings and schools and, therefore, should not be regarded as outdated.

The specific aim of this study is to evaluate the intra- and inter-rater reliability of the modified versions of the pectoral and hip flexor muscle length tests, with measurements taken on the same day as well as after 2- and 5-day intervals in children aged 10 to 15 years. The results will help determine whether the modified versions of these tests can be effectively used in scientific and other settings.

MATERIAL AND METHODS

Design

This was a reliability study with repeated measurements of the two muscle length test performed by the two raters on the same day and with the two- and five-day intervals (Figure 1).

In order to see all comparisons taken into account, all narrow linear arrows (inter-rater reliability) connecting blocks of rater A and rater B should be mirrored to the opposite side. Similarly, all solid block arrows (intra-rater reliability) interconnecting consecutive rater B blocks should bereflected to the opposite side. Intra-rater reliability 2/5 = within-rater reliability of measurements performed with 2- and 5-day intervals; inter-rater reliability 0/2/5 = between-rater reliability of measurements performed on one day and with a 2 and 5-day interval, respectively.

Participants

Sixty-four children from three local primary schools volunteered to participate in the study, comprising 28 boys and 36 girls from years 4 to 8 (aged 10-15 years). The participants were generally healthy children with varying daily physical activity levels (from sedentary lifestyles to recreational and sports activities). The inclusion criteria were as follows: an age range of 10 to 15 years (to encompass the entire pubertal growth spurt), typical neurological and motor development (no medical diagnoses of illnesses or dysfunctions, body mass and height within the 25th to 75th centiles for their age), and the ability to follow verbal instructions. The exclusion criteria included: a lifetime history or current diagnoses of any serious orthopedic or neurological conditions (e.g., fractures, congenital deformities, cerebral palsy, musculoskeletal pain or dysfunction lasting more than two weeks), a history of surgical interventions, recent musculoskeletal pain or dysfunction within one month before the study, or minor illnesses on the day of measurement (e.g., colds or headaches). Two girls were excluded due to a history of fractures. 62 children (34 girls and 28 boys) met the criteria and qualified for the study. Their mean age was 12.15 years (range: 10-15 years), with a mean body mass of 49.57 kg (range: 28-74 kg) and a mean height of 156.92 cm (range: 137.5-177.5 cm). The children and their parents were provided detailed information about the study's objectives and procedures, and parents gave their informed consent. The measurements were conducted at a local academic institution in the Motion Analysis Laboratory, and no participants dropped out during the procedure. Ethical approval for the study was obtained from the institutional Biomedical Research Ethics Committee (approval number 18/2020).

The chosen age range was selected for several reasons. Firstly, it allowed coverage of the entire pubertal growth spurt, a period during which postural examinations, including the two muscle length tests under investigation, are commonly performed by clinicians. This is also the stage when many postural abnormalities are identified.

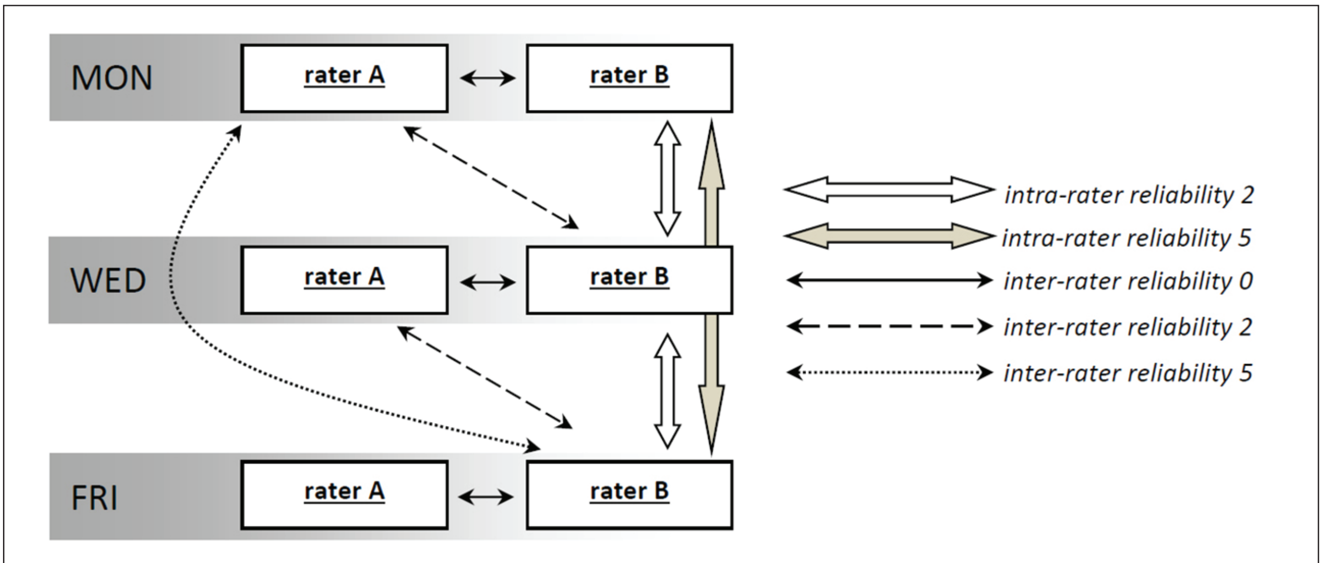

Figure 1. Simplified structure of the measurements and of calculation of the intraclass correlation coefficients.

Additionally, the variability in participants' posture served as a source of diversity, presenting challenges to the measurement process and enhancing the generalisability of the results. For the same reason, participants with differing daily physical activity levels were included.

Raters

Two raters with limited professional experience were included in the study to create challenging conditions for the research. Both raters were qualified physiotherapists with two years of clinical practice. They underwent an intensive two-week training program (three hours per day on weekdays) to develop adequate proficiency in performing the two muscle length tests. This training was supervised by an experienced specialist who was not directly involved in the study. Upon completion, the specialist determined that the raters were sufficiently skilled to perform their roles. The raters' abilities were further validated in a minor pilot study involving eight adult volunteers, during which the lowest recorded intra-class correlation coefficient (ICC) (model 3,3) was 0.93 for inter-rater reliability in the pectoral muscle length test performed on the same day.

An assistant was employed to read and record the measurement device outputs, which were concealed from the raters throughout the process to maintain blinding.

Pectoral muscles length test

During the measurement, participants assumed a relaxed supine position on the couch (Figure 2). Their heads were positioned slightly beyond the edge of the couch. At the same time, their legs were comfortably flexed (approximately 70 deg. at the hips and 120 deg. at the knees) with feet securely flat on the couch to relax the anterior structures of the lower trunk. Both wrists were placed through a strap loop (60 cm circumference), and participants were instructed to apply slight tension to the loop. This ensured a standard sagittal plane movement trajectory and minimised deviations into the frontal plane. To prevent thoracic spine extension, the lower ribcage was secured to the couch with a velcro strap positioned at the xiphoid process level. On the posterior aspect of each participant's arm, two points were marked 5 cm and 15 cm vertically above the olecranon process, measured in a standing position with the elbows extended. Two flat neodymium magnets (1.5 cm diameter) were affixed to these points using sticky tape, ensuring the distal edges of the magnets aligned with the skin markings. This setup enabled the quick assembly of the measurement device. The digital inclinometer Baseline-12-1057 (Baseline Products, USA) was selected for simplicity. This decision was informed by the findings of Moreside and McGill (2011), who observed no significant differences in reliability between inclinometric measurements and those using motion analysis systems. The inclinometer was calibrated in the horizontal position. The inclinometer's distal edge matched the distal magnet's distal edge. Both arms were prepared in this manner, with two inclinometers used simultaneously. The devices were positioned so their readouts faced the assistant, ensuring they were hidden from the rater.

With the wrist strap under slight tension, participants assumed the initial position: shoulders flexed to 90 deg., elbows fully extended, and thumbs extended and abducted, pointing cranially. The rater provided clear verbal instructions: "Inhale, exhale, let your arms smoothly lower behind your head." This movement had been

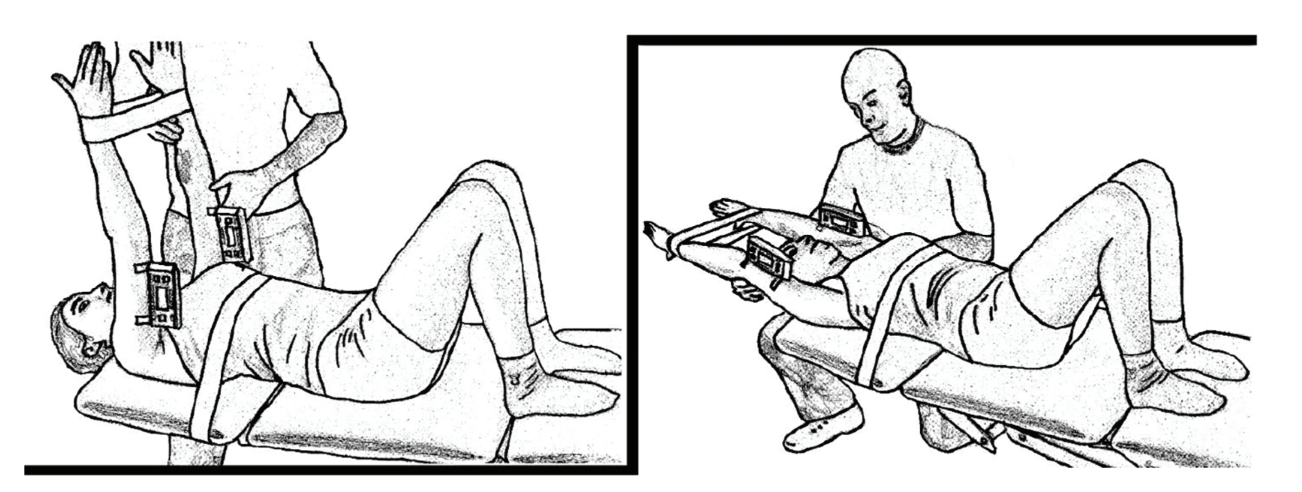

Figure 2. The modified pectoral muscles length test: initial position (left) and final position (right). To preserve clarity, the assistant is not present in the figure. The right- and left-side measurements were performed simultaneously.

https://doi.org/10.52291/ijse.2025.40.11

previously explained to participants during preparation. If necessary, the rater offered additional proprioceptive cues using their hands. Once the movement reached its endpoint and the participant's arms rested in the fully extended position, the rater issued the command "now" to signal the assistant to record the readings from both digital inclinometers.

Hip flexors length test

During the measurement, participants assumed a relaxed supine position on the couch, arms crossed over their chest and legs extended. On the anterior aspect of each participant's thigh, two points were marked: 5 cm and 15 cm proximally from the upper edge of the patella, measured with the knee joint fully extended. Using sticky tape, and two flat neodymium magnets (1.5 cm diameter) were affixed to these points, ensuring their distal edges aligned with the skin markings. This setup facilitated the quick assembly of the digital inclinometer (as described previously), which was calibrated in its horizontal position. The distal edge of the inclinometer was aligned with the distal edge of the distal magnet. Both lower limbs were prepared this way, with two inclinometers used simultaneously. The readouts of the inclinometers were directed outward from the midline of the participant's body, enabling the assistant to view the readout of the device on the measured side, while the rater could see the non-measured side.

Initially, the rater positioned the participant's legs in a flexed position, with the hip joint at 130 deg. of flexion and the knee in full available flexion (Figure 3). The 130 deg. hip flexion ensured the lumbar spine rested flat against the couch's surface, eliminating lordosis. The rater controlled the flexion range in the non-tested hip by monitoring the inclinometer's readout, which was visible to them. Clear verbal instructions were provided: "Inhale, exhale, let your leg smoothly lower down." This movement had been explained in detail to participants beforehand. During the movement, the rater manually stabilized the non-tested thigh at 130° of hip flexion and the pelvis near the anterior superior iliac spine on the tested side. Once the tested leg reached its final resting position, the rater gave the command "now" to the assistant, signalling them to record the digital inclinometer's reading on the tested thigh, which remained hidden from the rater.

After completing the measurement on one side, the rater and assistant switched positions at the participant's sides, and the roles of the lower limbs were reversed. The procedure was then repeated for the opposite side of the body.

Procedure

After completing the initial interview and verifying the selection criteria, parents and children selected a convenient week for the measurements. Participants arrived at the laboratory in the afternoon, ensuring they had at least three hours of rest after school and had eaten their last meal at least two hours beforehand. Children were instructed to wear loose, non-restrictive clothing and completed a 10-minute low-intensity warm-up on a stationary bike. Participants were then briefed on how to cooperate during the tests, and two to three practice attempts of each test were introduced. These initial attempts did not involve exploring the full range of motion. The order of the tests for each participant

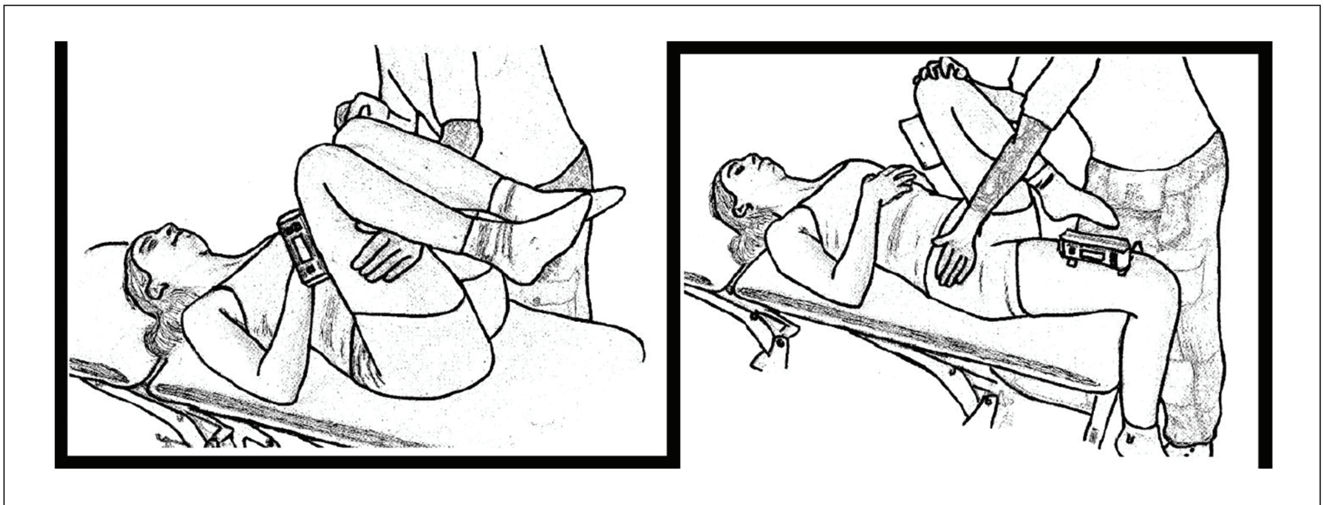

Figure 3. The modified hip flexors muscles length test: initial position (left) and final position (right). To preserve clarity the assistant is not present in the figure. The right-side measurement is presented. Subsequently, the procedure was repeated on the opposite side of the body.

was randomly determined on the first measurement day (Monday) and remained consistent across subsequent measurement days (Wednesday and Friday of the same week). For the hip flexors length test, the order was always the right side, followed by the left side. Raters were randomly assigned letters A and B, and the testing sequence always followed the order A-B. On Monday, after participants were appropriately prepared, Rater A conducted three repetitions of each muscle length test. All skin markings, magnets, and other equipment were then removed, and participants rested in a relaxed, semi-reclined position in an armchair for 15 minutes. Subsequently, Rater B prepared the participants for their measurements, and the two tests were repeated three times each. This procedure was identically repeated on Wednesday and Friday (Figure 1).

The measurements took place between March and June, ensuring a consistent functional context for the study, including a similar level of school-related activities, programs, and free time. Participants were instructed to maintain their usual levels of physical activity throughout the measurement week and avoid any sudden changes to their routines.

Data processing

Data recorded by the two raters was gathered in the Statistica 13 (Statistica, Tulsa, USA) spreadsheet. To calculate reliability indices for different types of reliability (Figure 1) and to avoid redundancy of results, this database was specifically restructured. Planned comparisons were pooled together in accordance with the following scheme:

- intra-rater reliability, the two-day interval between consecutive measurements: rater A on Mon *vs* rater A on Wed + rater A on Wed *vs* rater A on Fri + rater B on Mon *vs* rater B on Wed + rater B Wed *vs* rater B Fri; 248 records in total;
- intra-rater reliability, the five-day interval between consecutive measurements: rater A on Mon *vs* rater A on Fri + rater B on Mon *vs* rater B on Fri; 124 records in total;
- inter-rater reliability, measurements were taken on the same day: rater A on Mon *vs* rater B on Mon + rater A on Wed *vs* rater on B Wed + rater A on Fri *vs* rater B on Fri; 186 records in total;
- inter-rater reliability, the two-day interval between consecutive measurements: rater A on Mon *vs* rater B on Wed + rater A on Wed *vs* rater B on Fri + rater B on Mon *vs* rater A on Wed + rater B on Wed *vs* rater A on Fri; 248 records in total;
- inter-rater reliability, the five-day interval between consecutive measurements: rater A on Mon *vs* rater B on Fri + rater B on Mon *vs* rater A on Fri; 124 records in total.

Statistical analysis

A mixed-model ANOVA was employed to compare the mean values of the muscle length test results obtained by the two raters across consecutive measurement days, with the rater and side of the body as independent factors and measurement day as a repeated factor. A mixed-model ANOVA was applied to calculate intra-class correlation coefficients (ICCs), treating the results of consecutive measurements as a repeated factor and participants as an independent factor. The reliability of measurements taken on the same day by a single rater is not presented, as it is inherently encompassed within other forms of intra-rater reliability. Refer to Figure 1 for further details. Model 2, *k* of the ICC was applied to allow for generalisation to the broader population of raters with similar characteristics. ICCs were calculated separately for single repetitions of the muscle length tests, the mean value of two repetitions, and the mean value of three repetitions. Additionally, the standard errors of measurement (SEM = SD × (1–ICC) ^{½}) and the smallest detectable differences (SDD = 1.96 × $2^{½}$ × SEM) were computed. The interpretation of ICC values was as follows: 0.00-0.50 = poor reliability, 0.50-0.75 = moderate reliability, 0.75-0.90 = good reliability, and 0.90-1.00 = excellent reliability (Portney & Watkins, 2009).

RESULTS

The Table 1 presents the descriptive statistics for the results of the two muscle length tests of interest obtained by the two raters on consecutive measurement days. Given that the normative range for the pectoral muscle length test is 160-180 deg. and for the hip flexor length test is 0 deg. (Kendall et al., 2005; Peeler & Anderson, 2008), it can be concluded that the study group exhibited no more than mild shortenings. No significant differences were found for any of the main effects of the ANOVA (measurement day, rater, or side of the body), nor their interactions (all $P<0.05$).

Mean values ± standard deviations (minimum-maximum) based on the three repeated measurements obtained by the two raters on three consecutive measurement days.

The results of the two muscle length tests demonstrated good to excellent reliability, regardless of whether they

Table 1. **Mean values ± standard deviations (minimum-maximum) based on the three repeated measurements obtained by the two raters on three consecutive measurement days.**

		Pectoral muscles		Hip flexors	
	Rater	Left	Right	Left	Right
Monday	A	178.28 ± 9.68 (160.33-194.67)	181.75 ± 7.66 (167.33-199.33)	5.92 ± 4.76 (0.67-18.00)	7.11 ± 4.88 (1.33-19.67)
	B	176.91 ± 8.61 (161.67-192.16)	182.68 ± 8.11 (165.33-201.33)	6.08 ± 4.59 (1.33-18.67)	6.97 ± 4.65 (0.67-18.67)
Wednesday	A	179.06 ± 7.95 (162.33-193.33)	183.84 ± 8.71 (168.67-202.67)	6.12 ± 4.19 (1.67-16.67)	7.39 ± 4.58 (1.67-19.33)
	B	178.31 ± 9.11 (160.67-195.11)	181.29 ± 7.89 (164.33-200.33)	6.24 ± 3.99 (2.33-17.67)	7.45 ± 4.17 (2.33-19.00)
Friday	A	180.23 ± 8.64 (162.33-196.33)	179.91 ± 8.25 (165.67-197.67)	7.03 ± 5.01 (1.33-19.33)	7.94 ± 4.96 (1.67-20.33)
	B	179.57 ± 8.95 (162.33-194.78)	180.96 ± 8.56 (163.67-200.67)	6.85 ± 4.95 (2.33-18.33)	7.84 ± 4.66 (1.33-19.33)

Differences between measurement days, raters, and body sides were all non-significant (P<0.05).

Table 2. **Estimates of intra-rater reliability.**

	Reliability	**N meas.**	**Left**			**Right**		
			ICC (±95%CI)	SEM	SDD	ICC (±95%CI)	SEM	SDD
Pectoral muscles	intra-rater 2	1	0.92 (0.89-0.95)	2.25	6.25	0.96 (0.93-0.98)	1.77	4.92
		2	0.97 (0.95-0.98)	1.38	3.85	0.98 (0.96-0.99)	1.32	3.66
		3	0.98 (0.96-1.00)	1.23	3.41	0.98 (0.97-0.99)	1.16	3.21
	intra-rater 5	1	0.90 (0.87-0.92)	2.59	7.12	0.91 (0.89-0.94)	2.28	6.32
		2	0.93 (0.91-0.96)	1.73	4.79	0.95 (0.93-0.96)	1.65	4.57
		3	0.97 (0.95-0.99)	1.51	4.18	0.97 (0.95-0.99)	1.42	3.94
Hip flexors	intra-rater 2	1	0.88 (0.86-0.90)	1.63	4.51	0.86 (0.84-0.89)	1.97	5.47
		2	0.95 (0.94-0.97)	1.02	2.83	0.94 (0.92-0.96)	1.26	3.49
		3	0.97 (0.95-0.98)	0.79	2.19	0.96 (0.94-0.97)	0.98	2.72
	intra-rater 5	1	0.81 (0.75-0.92)	2.12	5.88	0.84 (0.75-0.92)	2.03	5.63
		2	0.88 (0.93-0.98)	1.74	4.82	0.93 (0.93-0.98)	1.29	3.58
		3	0.95 (0.95-0.99)	1.15	3.18	0.96 (0.95-0.99)	0.93	2.58

Presented are: intraclass correlation coefficient (ICC (±95% confidence interval); models: 2,1 for single measurements, 2,2 for two repeated measurements, and 2,3 for three repeated measurements), standard error of measurement (SEM (in degrees)) and smallest detectable difference (SDD (in degrees)). Intra-rater reliability 2 describes the reliability of measurements performed with a two-day interval (i.e., Monday-Wednesday, Wednesday-Friday); intra-rater reliability 5 describes the reliability of measurements performed with a five-day interval (i.e., Monday -Friday).

Table 3. **Estimates of inter-rater reliability.**

	Reliability	N meas.	Left			Right		
			ICC (±95%CI)	SEM	SDD	ICC (±95%CI)	SEM	SDD
Pectoral muscles	intra-rater 0	1	0.93 (0.89-0.95)	2.40	6.65	0.92 (0.89-0.94)	2.25	6.24
		2	0.96 (0.95-0.98)	1.83	5.07	0.95 (0.93-0.97)	1.78	4.93
		3	0.98 (0.96-1.00)	1.25	3.46	0.96 (0.95-0.98)	1.64	4.55
	intra-rater 2	1	0.88 (0.86-0.93)	2.98	8.26	0.89 (0.87-0.92)	2.75	7.62
		2	0.91 (0.89-0.94)	2.62	7.26	0.92 (0.90-0.95)	2.36	6.54
		3	0.95 (0.94-0.97)	1.97	5.46	0.94 (0.93-0.96)	2.01	5.57
	intra-rater 5	1	0.76 (0.73-0.79)	4.25	11.78	0.80 (0.77-0.84)	3.61	10.01
		2	0.85 (0.82-0.88)	3.32	9.20	0.87 (0.84-0.90)	2.91	8.07
		3	0.92 (0.90-0.94)	2.49	6.90	0.93 (0.91-0.94)	2.17	6.01
Hip flexors	intra-rater 0	1	0.91 (0.89-0.93)	1.42	3.94	0.90 (0.87-0.92)	1.43	3.96
		2	0.95 (0.93-0.97)	1.05	2.91	0.96 (0.94-0.98)	0.95	2.63
		3	0.97 (0.96-0.98)	0.79	2.19	0.98 (0.96-1.00)	0.66	1.83
	intra-rater 2	1	0.86 (0.83-0.88)	1.78	4.93	0.89 (0.86-0.92)	1.58	4.38
		2	0.92 (0.90-0.94)	1.33	3.69	0.94 (0.92-0.96)	1.19	3.30
		3	0.96 (0.95-0.98)	0.92	2.55	0.97 (0.96-0.99)	0.80	2.22
	intra-rater 5	1	0.82 (0.79-0.85)	1.89	5.24	0.85 (0.81-0.88)	1.88	5.21
		2	0.89 (0.86-0.91)	1.56	4.32	0.89 (0.86-0.91)	1.51	4.18
		3	0.91 (0.89-0.93)	1.37	3.80	0.94 (0.93-0.96)	1.14	3.16

Presented are: intraclass correlation coefficient (ICC (±95% Confidence Interval); models: 2,1 for single measurements, 2,2 for two repeated measurements, and 2,3 for three repeated measurements), standard error of measurement (SEM (in degrees)) and smallest detectable difference (SDD (in degrees)). Inter-rater reliability 0 describes the reliability of measurements performed on the same day; inter-rater reliability 2 describes the reliability of measurements performed with a two-day interval (i.e., Monday-Wednesday, Wednesday-Friday); inter-rater reliability 5 describes the reliability of measurements performed with a five-day interval (i.e., Monday -Friday).

were conducted by a single rater with a two- or five-day interval (Table 2) or by different raters on the same day or with a two- or five-day interval (Table 3). The lowest ICC recorded was 0.76, observed for the pectoral muscles length test on the left side of the body (SEM = 4.25, SDD = 11.78 deg.; inter-rater reliability based on a single repeated measurement taken with a five-day interval). In all other instances, ICCs of 0.80 or higher were achieved.

For intra-rater reliability (Table 2), even a single repeated measurement yielded ICCs exceeding 0.90. Increasing the number of repeated measurements led to a further rise in ICC values, with three repeated measurements resulting in ICCs of 0.95 or higher. A corresponding, gradual reduction in SEM and SDD values was also observed.

For inter-rater reliability (Table 3), lower ICC values were observed, particularly with longer intervals between measurements. In several cases, a single repeated measurement yielded ICCs below 0.90, with one instance (previously mentioned in this section) falling below 0.80. However, increasing the number of repeated measurements led to a significant improvement in ICC values. With three repeated measurements, all ICCs – including those calculated for measurements taken with a five-day interval – exceeded 0.91. As expected, all SEM and SDD values showed a consistent, gradual decrease.

DISCUSSION

Based on the results, we can conclude that the two muscle length tests under investigation achieved a satisfactory level of reliability. This level of reliability supports the use of the proposed modifications in scientific settings, offering a valuable tool for researchers interested in incorporating clinically oriented measurements into their studies. Such an approach is particularly beneficial in scenarios where advanced technological solutions may be overly specialized or restrictive. Furthermore, these modified and standardized tests could also prove useful in other settings, such as at home or in schools, and could be effectively utilized by a variety of professionals, including teachers, trainers, pedagogues, etc.

In our results, we chose to present the ICC model 2,*k* to enable generalization to broader populations of similar raters. Naturally, the ICC model 3,*k*, which reflects the reliability of measurements performed by a specific rater or group of raters, would yield higher values. Additionally, we did not report intra-rater reliability for measurements taken on the same day (referred to as intra-rater 0 in our nomenclature), as this is inherently embedded in the higher-order reliability models (intra-rater 2, 5; inter-rater 0, 2, 5). Logically, its ICC values would be higher due to the limited sources of variance.

To provide practical recommendations for using the two muscle length tests, we suggest that, in a scientific setting, it is preferable to perform three repeated measurements for each test. This is particularly important in cases where the time interval between test and retest is longer (up to five days in this study) or when different raters conduct the test and retest. In other circumstances, such as with a shorter time interval or when a single rater is involved, even two repeated measurements appear sufficient. The measurements, however, are not time-consuming, and adding one more repetition, 'just in case' would not be an expensive investment. We do not recommend reducing the number of repeated measurements to fewer than two. While such a reduction might be acceptable in a clinical, home, or school setting – provided the clinician is well-trained and strictly adheres to the procedure – scientific applications require higher reliability. For instance, with only one repeated measurement, the lowest ICC recorded in our study was 0.76 (pectoral muscles length test, inter-rater reliability for a five-day interval). While this is a reasonably satisfactory result for clinical or non-scientific settings, it does not meet the higher standards expected in research contexts.

It is worth noting that we successfully achieved our objectives under relatively challenging conditions. Firstly, we implemented a long interval between the test and retest sessions, with a maximum gap of five days. Secondly, we involved two raters in the procedures. Thirdly, these raters were relatively inexperienced and required specific training to perform the tests. Lastly, the age range of our participants encompassed the entire puberty growth spurt – a period characterized by dynamic morphological and functional changes in the human body. Each of these factors could potentially diminish the reliability of any measurement. Despite these challenges, we achieved ICC values indicating excellent reliability for the proposed modifications of the two muscle length tests. These findings strongly support their use in scientific research as well as in clinical, home, or school settings.

Unfortunately, there is limited literature available to directly compare our findings with. For the pectoral muscles length test, to the best of our knowledge, no studies have been conducted on its reliability. In the case of the hip flexors length test, Kim and Ha (2015) report similar levels of reliability, including SEM and SDD, and highlight that either active or passive pelvic stabilisation during the test improves reliability. However, their study involved a small sample size (n = 13) and participants who were older than those in our study (mean age 24.0 ± 3.9 years). Vigotsky et al. (2016) also emphasize the importance of pelvic stabilization, identifying it as a critical factor for achieving acceptable reliability in this test. Their results were based on a sample of 29 subjects aged 22.0 ± 3.8 years. In our study, we ensured adequate pelvic position control through passive manual stabilization and by maintaining a constant position of the contralateral hip. These measures likely contributed to the high reliability we achieved. Cady et al. (2022) documented moderate to high inter-rater reliability for a modified hip flexors length test; however, their results are difficult to compare with ours, as their analysis relied on evaluations of digital images rather than real-world measurements. In contrast, Peeler and Anderson (2008) were unable to achieve satisfactory reliability in their investigation, reporting ICCs of 0.67 (intra-rater) and 0.50 (inter-rater). It appears that their study lacked appropriate pelvic stabilization, used a less precise measurement method (a two-arm goniometer), and relied solely on single measurements without incorporating repeated measures. These factors likely contributed to their lower reliability scores.

LIMITATIONS

The primary limitation of our study lies in its restricted external validity, as the sample was limited to teenagers. This focus was chosen because we aimed to assess the reliability of the tests within a population where they are commonly used as part of postural examinations. However, caution is advised when attempting to generalize these findings to broader populations. The advantages of our modifications include minimal time, personnel, and equipment requirements. The procedure can be conducted by a single rater. The assistant in this study was employed solely for blinding purposes. Additionally, the equipment costs were kept low, requiring only two digital goniometers, in contrast to the more sophisticated motion analysis systems used in studies such as Kim and Ha (2015) and Vigotsky et al. (2016).

CONCLUSIONS

With appropriate specific modifications, it is possible to achieve very good to excellent levels of reliability for the pectoral and hip flexor muscle length tests. When three repeated measurements were utilized, all recorded ICCs exceeded 0.90. This high level of reliability supports the use of the modified versions of these tests in scientific settings. Additionally, these standardized and simplified tests could prove beneficial in other contexts, such as at home or in schools, and could be effectively employed by a range of professionals, including teachers, trainers, and pedagogues.

ACKNOWLEDGEMENT

None

DECLARATION OF INTEREST STATEMENT

The author reported no potential conflict of interest

FUNDING

None

REFERENCES

Barczyk, K., Zawadzka, D., Hawrylak, A., Bocheńska, A., Skolimowska, B., & Małachowska-Sobieska, M. (2009). The influence of corrective exercises in a water environment on the shape of the antero-posterior curves of the spine and on the functional status of the locomotor system in children with Io scoliosis. *Ortopedia Traumatologia Rehabilitacja*, *11*(3), 209-221.

Biały, M., Adamczyk, W., Gnat, R., & Stranc, T. (2017). Tissue Deformation Index as a Reliable Measure of Lateral Abdominal Muscle Activation on M-Mode Sonography. *Journal of Ultrasound in Medicine*, *36*(7), 1461-1467. https://doi.org/10.7863/ultra.16.07045

Biały, M., Adamczyk, W., Marczykowski, P., Majchrzak, R., & Gnat, R. (2019). Deformations of abdominal muscles under experimentally induced low back pain. *European Spine Journal*, *28*(11), 2444-2451. https://doi.org/10.1007/s00586-019-06016-y

Biały, M., Klaptocz, P., & Gnat, R. (2010). Functional asymmetry of the spine in standing and sitting positions. *Acta Gymnica*, *40*(1), 53-60.

Boguszewski, D., Adamczyk, J.G., Buda, M., & Kłoda, M. (2016). The use of functional tests to assess risk of injuries in judokas. *Archives of Budo Science of Martial Arts and Extreme Sports*, *12*, 57-62.

Cady, K., Powis, M., & Hopgood, K. (2022). Intrarater and interrater reliability of the modified Thomas Test. *Journal of Bodywork and Movement Therapies*, *29*, 86-91. https://doi.org/10.1016/j.jbmt.2021.09.014

Chmielewska, D., Cebula, M., Gnat, R., Rudek-Zeprzałka, M., Gruszczyńska, K., Baron, J., & Opala-Berdzik, A. (2023) Reliability of inter-recti distance measurement on ultrasound images captured by novice examiners. *Physiotherapy Theory and Practice*, *11*, 1-9. https://doi.org/10.1080/09593985.2023.2255897

D'Silva, P. V., Bandekar, P., Bhoir, H., & D'Silva, S. (2023). Prevalence of middle crossed syndrome in young healthy individuals from a health science university in India. *Journal of Bodywork and Movement Therapies*, *36*, 270-273. https://doi.org/10.1016/j.jbmt.2023.04.090

Dębski, P., Białas, E., & Gnat, R. (2019). The parameters of foam rolling, self-myofascial release treatment: a review of the literature. *Biomedical Human Kinetics*, *11*(1), 36-46. https://doi.org/10.2478/bhk-2019-0005

Gillani, S.N., Ain, Q., Rehman, S.U., & Masood, T. (2020). Effects of eccentric muscle energy technique versus static stretching exercises in the management of cervical dysfunction in upper cross syndrome: a randomized control trial. *The Journal of the Pakistan Medical Association*, *70*(3), 394-398. https://doi.org/10.5455/jpma.300417

Gnat, R., Biały, M., & Dziewońska, A. (2021). Experimentally induced low back pain influences brain networks activity. *Journal of Motor Behavior*, *53*(6), 680-692. https://doi.org/10.1080/00222895.2020.1839376

Gnat, R., & Biały, M. (2015). A new approach to the measurement of pelvic asymmetry: proposed methods and reliability. *Journal of manipulative and physiological therapeutics*, *38*(4), 295-301. https://doi.org/10.1016/j.jmpt.2015.02.002

Gnat, R., Dziewońska, A., Biały, M., & Wieczorek, M. (2020). Differences in Activity of the Brain Networks During Voluntary Motor Tasks Engaging the Local and Global Muscular Systems of the Lower Trunk. *Motor Control*, *24*(4), 624-643. https://doi.org/10.1123/mc.2019-0109

Gnat, R., Saulicz, E., & Miądowicz, B. (2012). Reliability of real-time ultrasound measurement of transversus abdominis thickness in healthy trained subjects. *European Spine Journal, 21*(8), 1508 -1515. https://doi.org/10.1007/s00586-012-2184-4

Gnat, R., Spoor, K., & Pool-Goudzwaard, A. (2015). The influence of simulated transversus abdominis muscle force on sacroiliac joint flexibility during asymmetric moment application to the pelvis. *Clinical biomechanics*, *30*(8), 827-831. https://doi.org/10.1016/j.clinbiomech.2015.06.006

Gogola, A., Gnat, R., Dziub, D., Gwóźdź, M., & Zaborowska, M. (2016). The impact of the neurodevelopmental traction technique on activation of lateral abdominal muscles in children aged 11–13 years. *NeuroRehabilitation*, *39*, 183-190. https://doi.org/10.3233/nre-161350

Gogola, A., Gnat, R., Zaborowska, M., Dziub, D., & Gwóźdź, M. (2018). Asymmetry of activation of lateral abdominal muscles during the neurodevelopmental traction technique. *Journal of Bodywork and Movement Therapies*, *22*(1), 46-51. https://doi.org/10.1016/j.jbmt.2017.03.019

Guruhan, S., Kafa, N., Ecemis, Z. B., & Guzel, N. A. (2021). Muscle Activation Differences During Eccentric Hamstring Exercises. *Sports Health*, *13*(2), 181-186. https://doi.org/10.1177/1941738120938649

Hecht, G. G., Van Rysselberghe, N. L., Young, J. L., & Gardner, M. J. (2022). Gait Analysis in Orthopaedic Surgery: History, Limitations, and Future Directions. *The Journal of the American Academy of Orthopaedic Surgeons*, *30*(21), e1366-e1373. https://doi.org/10.5435/jaaos-d-21-00785

Kendall, F. P., McCreary, E. K., Provance, P. G., Rodgers, M. M., & Romani, W. A. (2005). Muscles: Testing and Function with Posture and Pain. 5th ed. Lippincott Williams & Wilkins.

Kim, G. M., & Ha, S. M. (2015). Reliability of the modified Thomas test using a lumbo-plevic stabilization. *Journal of Physical Therapy Science*, *27*(2), 447-449. https://doi.org/10.1589/jpts.27.447

Kurashina, W., Iijima, Y., Sasanuma, H., Saito, T., & Takeshita, K. (2022). Evaluation of muscle stiffness in adhesive capsulitis with Myoton PRO. *JSES International*, *7*(1), 25-29. https://doi.org/10.1016/j.jseint.2022.08.017

Kurashina, W., Takahashi, T., Sasanuma, H., Saitsu, A., & Takeshita, K. (2023). Relationship Between Achilles Tendon Stiffness Using Myoton PRO and Translation Using a Tensile Testing Machine: A Biomechanical Study of a Porcine Model. *Curēus*, *15*(11), e49359. https://doi.org/10.7759/cureus.49359

Linek, P., Pałac, M., & Wolny, T. (2021). Shear wave elastography of the lateral abdominal muscles in C-shaped idiopathic scoliosis: a case-control study. *Scientific Reports*, *11*(1), 6026. https://doi.org/10.1038/s41598-021-85552-4

Moreside, J. M., & McGill, S. M. (2011). Quantifying normal 3D hip ROM in healthy young adult males with clinical and laboratory tools: Hip mobility restrictions appear to be plane-specific. *Clinical Biomechanics*, *8*(26), 824-829. https://doi.org/10.1016/j.clinbiomech.2011.03.015

Oliva-Lozano, J. M., & Muyor, J. M. (2020). Core Muscle Activity During Physical Fitness Exercises: A Systematic Review. *International Journal of Environmental Research and Public Health*, *17*(12), 4306. https://doi.org/10.3390/ijerph17124306

Ormachea, J., & Parker, K. J. (2020). Elastography imaging: the 30 year perspective. *Physics in Medicine and Biology*, *65*(24), 24TR06. https://doi.org/10.1088/1361-6560/abca00

Park, S. D. (2013). Reliability of Ultrasound Imaging of the Transversus Deep Abdominial, Internal Oblique and External Oblique Muscles of Patients with Low Back Pain Performing the Drawing-in Maneuver. *Journal of Physical Therapy Science*, *25*, 845-847. https://doi.org/10.1589/jpts.25.845

Peeler, J. D., & Anderson, J. E. (2007). Reliability of the Thomas Test for assessing range of motion about the hip. *Physical Therapy in Sport*, *8*(1), 14-21. https://doi.org/10.1016/j.ptsp.2006.09.026

Peeler, J. D., & Anderson, J. E. (2008). Reliability limits of the modified Thomas test for assessing rectus femoris muscle flexibility about the knee joint. *Journal of Athletic Training*, *43*(5), 470-476. https://doi.org/10.4085/1062-6050-43.5.470

Portney, L., & Watkins, M. (2009). *Foundations of Clinical Research: Applications to Practice*. 3rd ed. Pearson Education.

Qiu, Y., Guan, Y., & Liu, S. (2023). The analysis of infrared high-speed motion capture system on motion aesthetics of aerobics athletes under biomechanics analysis. *Public Library of Science one*, *18*(5), e0286313. https://doi.org/10.1371/journal.pone.0286313

Vigotsky, A. D., Lehman, G. J., Beardsley, C., Contreras, B., Chung, B., & Feser, E. H. (2016). The modified Thomas test is not a valid measure of hip extension unless pelvic tilt is controlled. *PeerJ*, *4*, e2325. https://doi.org/10.7717/peerj.2325